

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 08:30:38 ; Search time 450.858 seconds
(without alignments)
6388.441 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggcctctctcgagaagcgt.....gccaccacctgttctctgttaa 678

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678	100.0	678	6	AAD46282
2	676.4	99.8	678	6	AAD28209
3	670.2	98.8	675	6	AAD46281
4	668.4	98.6	678	6	AAD28208
5	665.2	98.1	678	6	AAD46278
6	665.2	98.1	678	6	AAD28207
7	665.2	98.1	678	6	AAA48743
8	661.6	97.6	681	9	ADC24134
9	661.6	97.6	4692	6	AAH47954
10	661.6	97.6	4692	7	ACC44640
11	661.6	97.6	7910	4	AAD09979
12	661.6	97.6	9320	6	ABS56684
13	660	97.3	5436	4	AAD10003
14	659.6	97.3	6893	9	ADE24111
15	655.2	96.6	723	7	ABZ22476
16	650.6	96.0	6985	7	ACA55359
17	649.8	95.8	678	9	ADC24128
18	629.6	92.9	681	9	ADC24130
19	612.6	90.4	681	4	AAH47656
20	568.6	83.9	678	9	ADC24132
21	556.4	82.1	678	3	AAA48748
22	556.4	82.1	678	6	AAD46280
23	553.2	81.6	678	3	AAA48749

24	551.6	81.4	678	3	AAA48750
25	460.4	67.9	678	6	ABA95905
26	460.4	67.9	678	6	ABA95920
27	459.2	67.7	681	6	ABA95906
28	459.2	67.7	681	6	ABA95907
29	438	64.6	678	4	AAH47654
30	438	64.6	678	4	AAD13053
31	438	64.6	678	4	AAD11142
32	438	64.6	678	6	ABA95921
33	438	64.6	678	9	ADC24127
34	438	64.6	859	6	AAH47952
35	438	64.6	859	7	AAD53432
36	438	64.6	859	9	AAD61969
37	437	64.5	3311	8	ACA62995
38	432	63.7	666	6	ABL61142
39	431.6	63.7	678	6	ABA95922
40	422	62.2	898	4	AAD03615
41	420.4	62.0	898	3	AAA48742
42	377.8	55.7	876	3	AAA28958
43	377.8	55.7	876	4	AAD03618
44	377.8	55.7	876	7	ACC44649
45	318.2	46.9	696	3	AAA50885

ALIGNMENTS

RESULT 1
AAD46282
ID AAD46282 standard; DNA; 678 BP.

AC AAD46282;
DT 27-DEC-2002 (first entry)

DE Discosoma sp. drFP583 (NFP-6) mutant DNA, E5-NA.

XX Fluorescent protein; chromoprotein; protease cleavage assay; filter;
XX fluorescence activated cell sorting application; fluorescent timer;
XX biosensor; fluorescence resonance energy transfer application; FRET;
XX colouring agent; recombinant DNA application; analyte detection assay;
XX sunscreen; second messenger detector; drFP583 protein; NFP-6; gene;
XX mutant; ds.

OS Discosoma sp.
OS Synthetic.

FX Key Location/Qualifiers

FT CDS 1..678
FT CDS /*tag= a /product= "drFP583 mutant protein, E5-NA"

FT WO200268459-A2.

PD 06-SEP-2002.

XX 20-FEB-2002; 2002WO-US005749.

XX AAL47954 Modified

PR Acc44640 Vector pD

PR AAD09979 pEIT(dHSP

XX ABS56684 Plasmid p

XX AAD10003 Plasmid p

PI ADE24111 Proviral

XX ABZ22476 Mammalian

XX ACA55359 Transform

PS Disclosure; Page 73; 80pp; English.

XX The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drfp583 (NFP-6) mutant DNA of the invention
 XX SQ Sequence 678 BP; 145 A; 225 C; 203 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 6; Length 678;
 Best Local Similarity 100.0%; Pred. No. 1.7e-109;
 Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGTTCAGGTGCGCATGGAGGGC 60
 Db 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGTTCAGGTGCGCATGGAGGGC 60

Qy 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGCGAGGGCGCCCTACGAGGGC 120
 Db 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGCGAGGGCGCCCTACGAGGGC 120

Qy 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGGGGCCCTCGCCCTCGCGCATC 180
 Db 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGGGGCCCTCGCCCTCGCGCATC 180

Qy 181 CTGTCCCTCCAGTTCAGTACGGCTCCAGGTGAGGTGAGTGAAGCACCCCGCCGATCCCC 240
 Db 181 CTGTCCCTCCAGTTCAGTACGGCTCCAGGTGAGGTGAGTGAAGCACCCCGCCGATCCCC 240

Qy 241 GACTACAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCTGATGAATCTCGAG 300
 Db 241 GACTACAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCTGATGAATCTCGAG 300

Qy 301 GACGGCGCTGGCGACCGGTGACCCAGGACTCTCTCTCGAGGACGGCTGCTTCATCTAC 360
 Db 301 GACGGCGCTGGCGACCGGTGACCCAGGACTCTCTCTCGAGGACGGCTGCTTCATCTAC 360

Qy 361 AAGTGAAGTTCATCGCGTGAATTCCTCTCGACGGCCCTGATGAGAGAGAC 420
 Db 361 AAGTGAAGTTCATCGCGTGAATTCCTCTCGACGGCCCTGATGAGAGAGAC 420

Qy 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGCTGCTGAAGGGCGAG 480
 Db 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGCTGCTGAAGGGCGAG 480

Qy 481 ATCCAAAGCCCTGAAGCTGAAGGACGGCGGCATCTACTGTGAGTTCAGTCCATC 540
 Db 481 ATCCAAAGCCCTGAAGCTGAAGGACGGCGGCATCTACTGTGAGTTCAGTCCATC 540

Qy 541 TACATGCCAAGAGCCCGTGCAGTGCCTGCTGCTGACACCAAGCTGGAC 600
 Db 541 TACATGCCAAGAGCCCGTGCAGTGCCTGCTGCTGACACCAAGCTGGAC 600

Qy 601 ATCACTCTCCCAACGAGGACTACACATCTGTGGAGCAGTACGAGCGCACCGAGGCC 660
 Db 601 ATCACTCTCCCAACGAGGACTACACATCTGTGGAGCAGTACGAGCGCACCGAGGCC 660

Qy 661 CACCACCTGTCTCTGTA 678
 Db 661 CACCACCTGTCTCTGTA 678

RESULT 2
 AAD28209 standard; cDNA; 678 BP.
 XX AAD28209;
 AC AAD28209;
 XX

DT 22-APR-2002 (first entry)
 XX Discosoma sp. non-aggregating mutant timer protein ESNA encoding cDNA.
 DE Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic plant; gene modification;
 KW protein age; mutant; ESNA; ss.
 XX Discosoma sp.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT 1..678
 CDS /tag= a
 /product= "Mutant ESNA protein"
 XX WO200196373-A2.
 XX 20-DEC-2001.
 PD 13-JUN-2001; 2001WO-US019097.
 PF 14-JUN-2000; 2000US-0211607P.
 PR (CLON-) CLONTECH LAB INC.
 XX Fradkov AF, Tersikh A;
 PI WPI: 2002-154595/20.
 DR P-PSDB; AAE17542.
 XX New fluorescent timer proteins comprising an emission spectrum that
 PT changes over time from a first wavelength to a second wavelength, useful
 PT for monitoring intracellular protein movement, translocation, trafficking
 PT or stability.
 XX Disclosure; Fig 3; 89pp; English.
 CC The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in
 CC enriching a population of cells comprising a fluorescent timer protein.
 CC The fluorescent timer proteins are also useful for assessing gene
 CC expression during development of a multicellular organism or during
 CC cellular differentiation, in response to a drug or other inducer of
 CC promoter activity, as a reporter to serve as a read-out of promoter
 CC activity, monitoring intracellular protein movement or translocation,
 CC protein trafficking, or protein stability, to investigate temporal
 CC aspects of the activity of a regulatory element, for determining cell
 CC fate during development and organ remodeling, in spatial and temporal
 CC visualisation of newly synthesised proteins and accumulated proteins, and
 CC in distinguishing between newly formed and pre-existing structures, e.g.
 CC membrane junctions and extracellular matrix components. The fluorescent
 CC timer proteins may further be used to investigate where photobleaching
 CC techniques are employed, as detectable labels, as selectable markers, as
 CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
 CC assays, and as second messenger detectors. The nucleic acids can be used
 CC to generate transgenic, non-human plants or animals or site-specific gene
 CC modifications in cell lines. The present sequence is a cDNA encoding
 CC Discosoma sp. ESNA non-aggregating fluorescent timer protein derived from
 CC fluorescent timer protein E5 by substituting amino acids at positions
 CC R2A, K5E and K9T. E5 protein is derived from humanised wild-type Anthozoa
 CC protein drfp583 by substituting Val to Ala at 105 and Ser to Thr at 197
 XX SQ Sequence 678 BP; 145 A; 226 C; 203 G; 104 T; 0 U; 0 Other;

Query Match 99.8%; Score 676.4; DB 6; Length 678;
 Best Local Similarity 99.9%; Pred. No. 3.2e-109;
 Matches 677; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGTTCAGGTGCGCATGGAGGGC 60

Db	1	ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGATGAGGGC	60
QY	61	ACCGTGAAGCGCCAGAGTTTCAGATCGAGGGCGAGGGCGCGCCCTACGAGGGC	120
Db	61	ACCGTGAAGCGCCAGAGTTTCAGATCGAGGGCGAGGGCGCGCCCTACGAGGGC	120
QY	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGCGCCCTTCGCTTGGGACATC	180
Db	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGCGCCCTTCGCTTGGGACATC	180
QY	181	CTGTCCCCCGGCTTCAGTTCAGGTCACGCTCCAGGTGTAGTGAGCAGCCCGCGACATCCC	240
Db	181	CTGTCCCCCGGCTTCAGTTCAGGTCACGCTCCAGGTGTAGTGAGCAGCCCGCGACATCCC	240
QY	241	GACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAATTCGAG	300
Db	241	GACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAATTCGAG	300
QY	301	GACGGCGGTGGGACCGTGACCCAGAGCTCTCTCCCTGCGAGGCGGTCTTCACTAC	360
Db	301	GACGGCGGTGGGACCGTGACCCAGAGCTCTCTCCCTGCGAGGCGGTCTTCACTAC	360
QY	361	AAGGTGAAGTTCATCGGCTGAACCTTCCCTCCGAGCGCCGCTGATGCAGAGAAGACC	420
Db	361	AAGGTGAAGTTCATCGGCTGAACCTTCCCTCCGAGCGCCGCTGATGCAGAGAAGACC	420
QY	421	ATGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGCGGTGCTGAAGGGCGAG	480
Db	421	ATGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGCGGTGCTGAAGGGCGAG	480
QY	481	ATCCACAAGCGCTGAAGCTGAAGAGCGGCGGCACTACTGCTGGAGTTCAAGTCCATC	540
Db	481	ATCCACAAGCGCTGAAGCTGAAGAGCGGCGGCACTACTGCTGGAGTTCAAGTCCATC	540
QY	541	TACATGCGCAAGAAGCGCGTGCAGCTGCGCGGCTACTACTAGTGGACACCAAGCTGGAC	600
Db	541	TACATGCGCAAGAAGCGCGTGCAGCTGCGCGGCTACTACTAGTGGACACCAAGCTGGAC	600
QY	601	ATCAGCTCCGACAGAGAGTACACCATCTGGAGCAGTACGAGGCGACCGAGGGCGGC	660
Db	601	ATCAGCTCCGACAGAGAGTACACCATCTGGAGCAGTACGAGGCGACCGAGGGCGGC	660
QY	661	CACCACTGTTCTCTGTAA 678	
Db	661	CACCACTGTTCTCTGTAA 678	
RESULT 3			
ID	AAD46281		
XX	AAD46281 standard; DNA; 675 BP.		
AC	AAD46281;		
XX			
DT	27-DEC-2002 (first entry)		
XX			
DE	Discosoma sp. drFP583 (NFP-6) mutant DNA, E57-NA.		
XX			
KW	Fluorescent protein; chromoprotein; protease cleavage assay; filter;		
KW	fluorescence activated cell sorting application; fluorescent timer;		
KW	biosensor; fluorescence resonance energy transfer application; FRET;		
KW	colouring agent; recombinant DNA application; analyte detection assay;		
KW	sunscreen; second messenger detector; drFP583 protein; NFP-6; gene;		
XX	mutant; ds.		
OS	Discosoma sp.		
OS	Synthetic.		
XX			
Key	Location/Qualifiers		
CDS	1..675		
FT	/*tag= a		
FT	/product= "drFP583 mutant protein, E57-NA"		
FT	/note= "No stop codon"		

FT	XX	/partial	
XX	XX	WO200268459-A2.	
XX	XX	06-SEP-2002.	
PD	XX		
XX	XX	20-FEB-2002; 2002WO-US005749.	
FF	XX		
XX	XX	21-FEB-2001; 2001US-0270983P.	
PR	XX	04-DEC-2001; 2001US-00006922.	
XX	XX	(CLON-) CLONTECH LAB INC.	
PA	XX	Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;	
FI	XX	WPI; 2002-691654/74.	
XX	XX	P-PSDB; AAE28836.	
XX	XX	New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant	
PT	XX	of an aggregating Chidarian chromo- or fluorescent protein or mutant for	
PT	XX	analyte detection assays or fluorescence activated cell sorting	
PT	XX	applications.	
XX	XX	Disclosure; Page 73; 80pp; English.	
PS	XX	The invention relates to nucleic acid molecules encoding non-aggregating	
CC	XX	chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are	
CC	XX	useful in analyte detection assays, as colouring agents, as markers in	
CC	XX	recombinant DNA applications, as sunscreens or filters, in fluorescence	
CC	XX	resonance energy transfer (FRET) applications, as biosensors in	
CC	XX	prokaryotic and eukaryotic cells, in screening assays, as second	
CC	XX	messenger detectors, in fluorescence activated cell sorting applications,	
CC	XX	in protease cleavage assays or as fluorescent timers. The present	
CC	XX	sequence is Discosoma sp. drFP583 (NFP-6) mutant DNA of the invention	
XX	XX	Sequence 675 BP; 142 A; 227 C; 204 G; 102 T; 0 U; 0 Other;	
SQ	XX	Query Match 98.8%; Score 670.2; DB 6; Length 675;	
	XX	Best Local Similarity 99.6%; Pred. No. 3.8e-108;	
	XX	Matches 672; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	1	ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGATGAGGGC	60
Db	1	ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGATGAGGGC	60
QY	61	ACCGTGAAGCGCCAGAGTTTCAGATCGAGGGCGAGGGCGCGCCCTACGAGGGC	120
Db	61	ACCGTGAAGCGCCAGAGTTTCAGATCGAGGGCGAGGGCGCGCCCTACGAGGGC	120
QY	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGCGCCCTTCGCTTGGGACATC	180
Db	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGCGCCCTTCGCTTGGGACATC	180
QY	181	CTGTCCCCCGGCTTCAGTTCAGGTCACGCTCCAGGTGTAGTGAGCAGCCCGCGACATCCC	240
Db	181	CTGTCCCCCGGCTTCAGTTCAGGTCACGCTCCAGGTGTAGTGAGCAGCCCGCGACATCCC	240
QY	241	GACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAATTCGAG	300
Db	241	GACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAATTCGAG	300
QY	301	GACGGCGGTGGGACCGTGACCCAGAGCTCTCTCCCTGCGAGGCGGTCTTCACTAC	360
Db	301	GACGGCGGTGGGACCGTGACCCAGAGCTCTCTCCCTGCGAGGCGGTCTTCACTAC	360
QY	361	AAGGTGAAGTTCATCGGCTGAACCTTCCCTCCGAGCGCCGCTGATGCAGAGAAGACC	420
Db	361	AAGGTGAAGTTCATCGGCTGAACCTTCCCTCCGAGCGCCGCTGATGCAGAGAAGACC	420
QY	421	ATGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGCGGTGCTGAAGGGCGAG	480
Db	421	ATGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGCGGTGCTGAAGGGCGAG	480

QY 481 ATCCACAGGCGCTCAAGCTGAAGAGCGGGCCACCTACCTGGTGGAGTTCAAGTCCATC 540
 DB |||||
 QY 481 ACCCACAGGCGCTGAAGCTGAAGAGCGGGCCACCTACCTGGTGGAGTTCAAGTCCATC 540
 DB |||||
 QY 541 TACATGGCCAAAGAGCGCGTGCAGCTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
 DB |||||
 QY 541 TACATGGCCAAAGAGCGCGTGCAGCTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
 DB |||||
 QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB |||||
 QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB |||||
 QY 661 CACCACCTGTTCCTG 675
 DB |||||
 QY 661 CACCACCTGTTCCTG 675
 DB |||||

RESULT 4

AD28208
 ID AAD28208 standard; cDNA; 678 BP.

XX AC AAD28208;
 XX AC

DT 22-APR-2002 (first entry)

XX Discosoma sp. fluorescent protein E5 encoding cDNA.

XX Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic plant; gene modification;
 KW protein age; E5; ss.

XX Discosoma sp.

XX Key Location/Qualifiers
 FH 1. 678
 CDS /*tag= a
 FT /product= "E5 protein"
 FT

XX WO200196373-A2.

XX 20-DEC-2001.

XX 13-JUN-2001; 2001WO-US019097.

XX 14-JUN-2000; 2000US-0211607P.

XX (CLON-) CLONTECH LAB INC.

XX Fradkov AF, Tersikh A;

XX WPI: 2002-154595/20.

XX P-PSDB; AAE17541.

XX New fluorescent timer proteins comprising an emission spectrum that
 PT changes over time from a first wavelength to a second wavelength, useful
 PT for monitoring intracellular protein movement, translocation, trafficking
 PT or stability.

XX Disclosure; Fig 2; 89pp; English.

XX The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in
 CC enriching a population of cells comprising a fluorescent timer protein.
 CC The fluorescent timer proteins are also useful for assessing gene
 CC expression during development of a multicellular organism or during
 CC cellular differentiation, in response to a drug or other inducer of
 CC promoter activity, as a reporter to serve as a read-out of promoter
 CC activity, monitoring intracellular protein movement or translocation,
 CC protein trafficking, or protein stability, to investigate temporal
 CC aspects of the activity of a regulatory element, for determining cell

CC fate during development and organ remodelling, in spatial and temporal
 CC visualisation of newly synthesised proteins and accumulated proteins, and
 CC in distinguishing between newly formed and pre-existing structures, e.g.
 CC membrane junctions and extracellular matrix components. The fluorescent
 CC timer proteins may further be used to investigate where photobleaching
 CC techniques are employed, as detectable labels, as selectable markers, as
 CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
 CC assays, and as second messenger detectors. The nucleic acids can be used
 CC to generate transgenic, non-human plants or animals or site-specific gene
 CC modifications in cell lines. The present sequence is a cDNA encoding
 CC Discosoma sp. E5 fluorescent protein derived from humanised wild-type
 CC Anthonia protein drFP583 by substituting Val to Ala at 105 and Ser to Thr
 CC at 197
 XX
 SQ Sequence 678 BP; 147 A; 224 C; 203 G; 104 T; 0 U; 0 Other;

Query Match 98.6%; Score 668.4; DB 6; Length 678;

Best Local Similarity 99.1%; Pred. No. 7.8e-108; Indels 0; Gaps 0;

Matches 672; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGC 60

DB |||||

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGC 60

DB |||||

QY 61 ACCGTGAAGCCGACGAGTTCGAGATCGAGGCGAGGCGGCGCCCTACGAGGC 120

DB |||||

QY 61 ACCGTGAAGCCGACGAGTTCGAGATCGAGGCGAGGCGGCGCCCTACGAGGC 120

DB |||||

QY 121 CACAACACCGTGAAGTGAAGTGCACCAAGGCGGCGCCCTCGCTTGGCGATC 180

DB |||||

QY 121 CACAACACCGTGAAGTGAAGTGCACCAAGGCGGCGCCCTCGCTTGGCGATC 180

DB |||||

QY 181 CTGTCCCTCCAGTTCAGATCGGCTCCAAAGTGTGAGGACCCCGCCGACATCCCC 240

DB |||||

QY 181 CTGTCCCTCCAGTTCAGATCGGCTCCAAAGTGTGAGGACCCCGCCGACATCCCC 240

DB |||||

QY 241 GACTACAGAAGCTGTCTTCCCGAGGCTTCAAGTGGAGCGCGTGATGAACCTCGAG 300

DB |||||

QY 241 GACTACAGAAGCTGTCTTCCCGAGGCTTCAAGTGGAGCGCGTGATGAACCTCGAG 300

DB |||||

QY 301 GACGCGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGCGCTCTCATCTAC 360

DB |||||

QY 301 GACGCGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGCGCTCTCATCTAC 360

DB |||||

QY 361 AAGGTGAAGTTCATCGGCGTGAAGTTCCTCCCGAGGCGCGCGTGATGCGAGAGAGACC 420

DB |||||

QY 361 AAGGTGAAGTTCATCGGCGTGAAGTTCCTCCCGAGGCGCGCGTGATGCGAGAGAGACC 420

DB |||||

QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGTGAAGGCGGAG 480

DB |||||

QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGTGAAGGCGGAG 480

DB |||||

QY 481 ATCCACAGGCGCTGAAGCTGAAGAGCGGCGCCACTACTGTGGAGTTCAAGTCCATC 540

DB |||||

QY 481 ATCCACAGGCGCTGAAGCTGAAGAGCGGCGCCACTACTGTGGAGTTCAAGTCCATC 540

DB |||||

QY 541 TACATGGCCAAAGAGCGCGTGCAGCTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600

DB |||||

QY 541 TACATGGCCAAAGAGCGCGTGCAGCTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600

DB |||||

QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660

DB |||||

QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660

DB |||||

QY 661 CACCACCTGTTCCTGTA 678

DB |||||

QY 661 CACCACCTGTTCCTGTA 678

DB |||||

RESULT 5

AAD46278

ID AAD46278 standard; DNA; 678 BP.

XX

PT for monitoring intracellular protein movement, translocation, trafficking
 XX or stability.

PS Example 1; Fig 1; 89pp; English.

XX The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in
 CC enriching a population of cells comprising a fluorescent timer protein.
 CC The fluorescent timer proteins are also useful for assessing gene
 CC expression during development of a multicellular organism or during
 CC cellular differentiation, in response to a drug or other inducer of
 CC promoter activity, as a reporter to serve as a read-out of promoter
 CC activity, monitoring intracellular protein movement or translocation,
 CC protein trafficking, or protein stability, to investigate temporal
 CC aspects of the activity of a regulatory element, for determining cell
 CC fate during development and organ remodeling, in spatial and temporal
 CC visualisation of newly synthesised proteins and accumulated proteins, and
 CC in distinguishing between newly formed and pre-existing structures, e.g.
 CC membrane junctions and extracellular matrix components. The fluorescent
 CC timer proteins may further be used to investigate where photobleaching
 CC techniques are employed, as detectable labels, as selectable markers, as
 CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
 CC assays, and as second messenger detectors. The nucleic acids can be used
 CC to generate transgenic, non-human plants or animals or site-specific gene
 CC modifications in cell lines. The present sequence is a DNA encoding
 CC Discosoma sp. humanised wild-type Anthozoa protein drFP583 used for
 CC generating fluorescent proteins

XX Sequence 678 BP; 146 A; 233 C; 203 G; 106 T; 0 U; 0 Other;

Query Match 98.1%; Score 665.2; DB 6; Length 678;

Best Local Similarity 98.8%; Pred. No. 2.8e-107;

Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGCTCTCCGAGACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGC 60

DB 1 ATGGCTCTCCGAGACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGC 60

QY 61 ACCGTGAACGCCACAGGTTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGCG 120

DB 61 ACCGTGAACGCCACAGGTTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGCG 120

QY 121 CACACACCGTGAAGTTGAAGTTGACCAAGGCGCGCCCTCGCTTGGCTGGGACATC 180

DB 121 CACACACCGTGAAGTTGAAGTTGACCAAGGCGCGCCCTCGCTTGGGACATC 180

QY 181 CTGTCCTCCAGTTCAGTACGCTCCAGAGTGTACGTGAAGCACCCCGCGACATCCCC 240

DB 181 CTGTCCTCCAGTTCAGTACGCTCCAGAGTGTACGTGAAGCACCCCGCGACATCCCC 240

QY 241 GACTACAAAGAGTGTCTTCCCGAGGCTTCAAGTGGAGCGCGTGAATCTCGAG 300

DB 241 GACTACAAAGAGTGTCTTCCCGAGGCTTCAAGTGGAGCGCGTGAATCTCGAG 300

QY 301 GAGGCGCGTGGAGCGCGTCCAGGACTCTCCCTCGAGACCGGCTTCATCTAC 360

DB 301 GAGGCGCGTGGAGCGCGTCCAGGACTCTCCCTCGAGACCGGCTTCATCTAC 360

QY 361 AAGTGAAGTTCATCGCGTGAATCTCCCTCCAGCGGCGCGTGTGAGAGAGAC 420

DB 361 AAGTGAAGTTCATCGCGTGAATCTCCCTCCAGCGGCGCGTGTGAGAGAGAC 420

QY 421 ATGGGCTGGAGCGCTTCCAGCGCGCTGTACCCCGGACCGCGCTGTGAGGCGGAG 480

DB 421 ATGGGCTGGAGCGCTTCCAGCGCGCTGTACCCCGGACCGCGCTGTGAGGCGGAG 480

QY 481 ATCCACAAGCGCTTGAAGTGAAGGACGCGGCGCACTACTGTGTGAGTTCAAGTCCATC 540

DB 481 ATCCACAAGCGCTTGAAGTGAAGGACGCGGCGCACTACTGTGTGAGTTCAAGTCCATC 540

QY 541 TACATGGCCAAAGCCCGTGCAGCTGCCCGCTACTACTACGTGGACACCAAGCTGGAC 600
 DB 541 TACATGGCCAAAGCCCGTGCAGCTGCCCGCTACTACTACGTGGACACCAAGCTGGAC 600
 QY 601 ATCACCTCCCAACAGGAGACTACACCATCTGTGGAGCGAGTACGAGCGGACCCGAGGGCGGC 660
 DB 601 ATCACCTCCCAACAGGAGACTACACCATCTGTGGAGCGAGTACGAGCGGACCCGAGGGCGGC 660
 QY 661 CACCACTGTTCTCTGTA 678
 DB 661 CACCACTGTTCTCTGTA 678

RESULT 7

AAA48743

ID AAA48743 standard; cDNA; 695 BP.

XX AAA48743;

AC AAA48743;

XX 19-SEP-2000 (first entry)

XX Humanised Discosoma sp. "red" novel fluorescent protein drFP583 cDNA.

XX Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;

KW fluorescent labeling; ss.

XX Discosoma sp; "red".

OS Synthetic.

XX WO200034326-A1.

XX 15-JUN-2000.

XX 10-DEC-1999; 99WO-US029473.

XX 11-DEC-1998; 98US-00210330.

PR 14-OCT-1999; 99US-00418529.

XX (CLON-) CLONTECH LAB INC.

PI Lukyanov SA, Fradkov AP, Labas YA, Matz MV, Green G, Chen Y;

PI Ding Li;

XX WPI; 2000-423381/36.

XX Novel fluorescent protein from non-bioluminescent Discosoma sp. red,

PT useful for fluorescent labeling and as markers.

XX Claim 6; Page 75-76; 86pp; English.

XX The present sequence is humanised drFP583 cDNA. drFP583 is a full-length
 CC cDNA encoding a novel fluorescent protein (nFP) from Discosoma sp. "red",
 CC a non-bioluminescent species of the Class Anthozoa. The wild-type drFP583
 CC nucleotide sequence was altered to optimise the codons for expression of
 CC the fluorescent protein in mammalian cells. Fluorescent proteins can be
 CC used in fluorescent labeling, a useful tool for marking a protein, cell
 CC or organism of interest. Unlike other markers used in protein labeling,
 CC such as beta-galactosidase and luciferase, fluorescent proteins do not
 CC require an exogenous cofactor or substrate. Methods involving fluorescent
 CC proteins are also less laborious and less difficult to control than the
 CC traditional methods of fluorescent labeling, where a protein of interest
 CC is purified and then covalently conjugated to a fluorophore derivative.
 CC Novel fluorescent proteins isolated from species of the Class Anthozoa
 CC can be used as markers for gene expression and protein localization
 CC studies, and in fluorescence resonance energy transfer (FRET) reactions.
 CC They may have improved properties and better suitability for larger
 CC fluorescent protein

XX Sequence 695 BP; 149 A; 228 C; 209 G; 109 T; 0 U; 0 Other;

Query Match 98.1%; Score 665.2; DB 3; Length 695;

Best Local Similarity 98.8%; Pred. No. 2.8e-107;

Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGGCTCTCCGAGAACTCATCCGAGTTTCATGCGCTTCAAGGTGGCATGAGGGC 60
DB |||||
QY 10 ATGGCTCTCCAGAACCTCATCAGAGTTTCATGCGCTTCAAGGTGGCATGAGGGC 69
DB |||||
QY 61 ACCGTGAACGGCCAGAGTTTCAGATTCGAGGGCGAGGGCGGGCCCTTACAGGGC 120
DB |||||
QY 70 ACCGTGAACGGCCAGAGTTTCAGATTCGAGGGCGAGGGCGGGCCCTTACAGGGC 129
DB |||||
QY 121 CACMACACCGTGAAGTTGAAGTTCACCAAGGGCGGGCCCTTCCCTTGGCTGGACATC 180
DB |||||
QY 130 CACMACACCGTGAAGTTCACCAAGGGCGGGCCCTTCCCTTGGCTGGACATC 189
DB |||||
QY 181 CTGTCCCTCCAGTTCCAGTACGGCTCCCAAGGTGTACGTGAAGCAACCCCGGACATCCC 240
DB |||||
QY 190 CTGTCCCTCCAGTTCCAGTACGGCTCCCAAGGTGTACGTGAAGCAACCCCGGACATCCC 249
DB |||||
QY 241 GACTACAAAGTGTCTTCCCGAGGGCTTCAAGTGGAGGGCGGTGATGAATCCGAG 300
DB |||||
QY 250 GACTACAAAGTGTCTTCCCGAGGGCTTCAAGTGGAGGGCGGTGATGAATCCGAG 309
DB |||||
QY 301 GACGGCGGGTGGCGACCGTGCACCCAGGACTCTCTCCCTGCAGGAGGGCTTCTATCTAC 360
DB |||||
QY 310 GACGGCGGGTGGCGACCGTGCACCCAGGACTCTCTCCCTGCAGGAGGGCTTCTATCTAC 369
DB |||||
QY 361 AGGTGAAGTTATCGCGGTGAATCTTCCCTCGAGGGCGGGCTGATGAGAGAGACC 420
DB |||||
QY 370 AGGTGAAGTTATCGCGGTGAATCTTCCCTCGAGGGCGGGCTGATGAGAGAGACC 429
DB |||||
QY 421 ATGGGTGGAGGGCTTCCAGAGGGCGGTGTACCCCGGAGGGCGGTGTGAGGGCGAG 480
DB |||||
QY 430 ATGGGTGGAGGGCTTCCAGAGGGCGGTGTACCCCGGAGGGCGGTGTGAGGGCGAG 489
DB |||||
QY 481 ATCCACAGCGCTGAAGTGAAGGACGGCGGCACTACTCGTGGAGTTCAAGTCCATC 540
DB |||||
QY 490 ATCCACAGCGCTGAAGTGAAGGACGGCGGCACTACTCGTGGAGTTCAAGTCCATC 549
DB |||||
QY 541 TACATGCCAAGAGCCCGTGCAGCTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
DB |||||
QY 550 TACATGCCAAGAGCCCGTGCAGCTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 609
DB |||||
QY 601 ATCACTCCCAACAGGAGCTACACATCGTGGAGCAGTACGAGGCGACCGAGGGCGCG 660
DB |||||
QY 610 ATCACTCCCAACAGGAGCTACACATCGTGGAGCAGTACGAGGCGACCGAGGGCGCG 669
DB |||||
QY 661 CACCACCTGTCTGTAA 678
DB |||||
QY 670 CACCACCTGTCTGTAA 687

RESULT 8
ADC24134
ID ADC24134 standard; DNA; 681 BP.
XX
AC ADC24134;
XX
AC
XX
DT 18-DEC-2003 (first entry)
XX
DE Discosoma wild-type red fluorescent protein DNA #2.
XX
KW Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
KW fluorescent protein variant; transcription induction detection;
KW fluorescence energy resonance transfer; FRET; protein kinase;
KW protein phosphatase; ion indicator; ds.
XX
OS Discosoma.
XX
PN US2003059835-A1.
XX
XX 27-MAR-2003.
XX
XX 10-APR-2002; 2002US-00121258.
PF

XX 26-FEB-2001; 2001US-00794308.
PR 24-MAY-2001; 2001US-00866538.
XX
XX (TSIE/) TSIE R Y.
PA (CAMP/) CAMPBELL R E.
PA
XX Tsien RY, Campbell RE;
PI WPI; 2003-743764/70.
XX P-PSDB; ADC24126.
DR Novel polynucleotide sequence encoding Discosoma red fluorescent protein
variant having a reduced propensity to oligomerize, useful for detecting
transcriptional activity.
XX
PS Disclosure; SEQ ID NO 9; 67bp; English.
XX
XX The invention describes a polynucleotide sequence (I) encoding a
Discosoma red fluorescent protein (DsRed) variant having a reduced
propensity to oligomerize, comprising amino acid substitutions at the AB
and/or AC interfaces of the wild-type DsRed sequence (81) comprising 225
amino acids, given in the specification, where the substitutions result
in reduced propensity of the DsRed variant to form tetramers. (I) is
useful for detecting transcriptional activity by providing a host cells
containing a vector which comprises (I) operatively linked to an
expression control sequence, and an unit to assay the variant fluorescent
protein fluorescence, and assaying fluorescence of the variant
protein fluorescent protein produced by (VII), where variant fluorescent protein
fluorescence is indicative of transcriptional activity. A polynucleotide
encoding a fusion protein is useful for the analysis of in vivo
localisation or trafficking of a polypeptide of interest. A polypeptide
marker is useful as markers to identify the location and amount of a
target protein produced, where the target protein is fused to the marker.
as a complement to or alternative for the green fluorescent protein or
its spectral variant, for detecting induction of transcriptions, in
applications involving fluorescence energy resonance transfer (FRET),
which detects events as the function of the movement of fluorescent
donors and acceptors towards or away from each other, for making
fluorescent sensors for protein kinase and phosphatase activities or
indicators for ions and molecules such as Ca²⁺, Zn²⁺, for identifying the
presence of a molecule in a sample, for identifying a specific
interaction of a first and second molecule, for determining whether a
sample contains an enzyme or for determining the pH of the sample. (I) is
useful for identifying a region or condition that regulates the activity
of an expression control sequence. This sequence represents a modified
Discosoma wild-type red fluorescent protein coding region with humanised
codon usage.
XX
SQ Sequence 681 BP; 146 A; 223 C; 205 G; 107 T; 0 U; 0 Other;
Query Match 97.6%; Score 661.6; DB 9; Length 681;
Best Local Similarity 98.7%; Pred. No. 1.2e-106;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGCTCTCTCGAGAACTCATCCGAGTTTCATGCGCTTCAAGGTGGCATGAGGGCA 61
DB |||||
QY 5 TGGCTCTCTCGAGAACTCATCCAGGAGTTTCATGCGCTTCAAGGTGGCATGAGGGCA 64
DB |||||
QY 62 CCGTGAACGGCCAGAGTTTCAGATTCGAGGGCGAGGGCGGGCCCTTACAGGGCC 121
DB |||||
QY 65 CCGTGAACGGCCAGAGTTTCAGATTCGAGGGCGAGGGCGGGCCCTTACAGGGCC 124
DB |||||
QY 122 ACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGGCCCTTCCCTTGGCTGGACATCC 181
DB |||||
QY 125 ACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGGCCCTTCCCTTGGCTGGACATCC 184
DB |||||
QY 182 TGTCCCCCGAGTTCAGTACCGCTCCAGGTGTACGTGAAGCACCCCGGACATCCCCG 241
DB |||||
QY 185 TGTCCCCCGAGTTCAGTACCGCTCCAGGTGTACGTGAAGCACCCCGGACATCCCCG 244
DB |||||
QY 242 ACTACAAGAGCTGTCTTCCCTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATTCGAGG 301
DB |||||

Db 245 ACTACAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 304
 QY 302 ACGGGGGGTGGACCGTGAACCCAGGACTCTCTCCCTGCGAGGACGGTGTCTTCACTACA 361
 Db 305 ACGGCGGGGTGGTACCGTGAACCCAGGACTCTCTCCCTGCGAGGACGGTGTCTTCACTACA 364
 QY 362 AGGTGAAGTTTCATCGGGGTGAACCTTCCCTCCGAGCGGCCCTGATGCAGAAAGACCA 421
 Db 365 AGGTGAAGTTTCATCGGGGTGAACCTTCCCTCCGAGCGGCCCTGATGCAGAAAGACCA 424
 QY 422 TGGGCTGGAGGCTCTCAACGAGGCGCTGTACCCCGCGGACGGGTGCTGAAGGGCGAGA 481
 Db 425 TGGGCTGGAGGCTCTCAACGAGGCGCTGTACCCCGCGGACGGGTGCTGAAGGGCGAGA 484
 QY 482 TCCACAAGGCGCTGAAGCTGAAGGAGCGGCGCACTACCTGTTGGAGTTTCAAGTCCATCT 541
 Db 485 TCCACAAGGCGCTGAAGCTGAAGGAGCGGCGCACTACCTGTTGGAGTTTCAAGTCCATCT 544
 QY 542 ACATGGCAAGAGCCCGTGCAGTGCCTCCGCTACTACTACTACTACTACTACTACTACTACT 601
 Db 545 ACATGGCAAGAGCCCGTGCAGTGCCTCCGCTACTACTACTACTACTACTACTACTACTACT 604
 QY 602 TCACCTCCCAACAGGAGCTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 661
 Db 605 TCACCTCCCAACAGGAGCTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 664
 QY 662 ACCACCTGTTCCCTGTA 677
 Db 665 ACCACCTGTTCCCTGTA 680

RESULT 9

AAL47954

ID AAL47954 standard; DNA; 4692 BP.

XX AC AAL47954;

XX DT 26-SEP-2002 (first entry)

XX DE Modified yeast strain related vector pDsRed1-N1.

XX DE Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;

XX DE modified yeast strain; environmental pollution; vector; ds.

XX OS Synthetic.

XX FN DE10061872-A1.

XX XX 20-JUN-2002.

XX PF 12-DEC-2000; 2000DE-01061872.

XX PR 12-DEC-2000; 2000DE-01061872.

XX PA (LICH/) LICHTENBERG-FRATE H.

XX PI Lichtenberg-Frate H;

XX DR WFI; 2002-539633/58.

XX Modified yeast strain, useful for detecting toxic compounds in
 XX environment, contains integrated cassettes responsive to genotoxic and
 XX cytotoxic compounds.
 XX Example 1; Page 24-25; 34pp; German.
 XX The present invention relates to a modified yeast strain that contains,
 XX integrated stably and functionally in its genome, a genotoxicity cassette
 XX and a cytotoxicity cassette, each comprising a promoter and reporter
 XX gene, both of which are different in the two cassettes. The modified
 XX yeast strain is used to detect environmental pollution, especially
 XX genotoxic and/or cytotoxic substances in complex environmental
 XX contaminants, especially organic compounds, but also (non-)ionising

CC radiation and chemical carcinogens. Particular applications are in
 CC monitoring (waste) water (e.g. as an early warning system), medical
 CC toxicology screening and for industrial process control. The present
 CC sequence is a vector suitable for use in the present invention
 XX
 SQ Sequence 4692 BP; 1111 A; 1268 C; 1247 G; 1066 T; 0 U; 0 Other;

Query Match 97.6%; Score 661.6; DB 6; Length 4692;
 Best Local Similarity 98.7%; Pred. No. 1.1e-106;
 Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGGCTCTCCGAGAAAGTTCATACCGAGTTTCATCGGCTTCAAGGTGCGATGAGGGCA 61
 Db 683 TGGGCTCTCCGAGAAAGTTCATACCGAGTTTCATCGGCTTCAAGGTGCGATGAGGGCA 742
 QY 62 CCGTGAACGGCCACGAGTTTCGAGATCGAGGCGAGGGCGGCGCCCTTACGAGGGCC 121
 Db 743 CCGTGAACGGCCACGAGTTTCGAGATCGAGGCGAGGGCGGCGCCCTTACGAGGGCC 802
 QY 122 ACAACACCGTGAAGTTGAAGTGAAGGTGACCAAGGCGGCGCCCTTCCCTTGGGACATCC 181
 Db 803 ACAACACCGTGAAGTTGAAGTGAAGGTGACCAAGGCGGCGCCCTTCCCTTGGGACATCC 862
 QY 182 TGTCCCCCAGTTTCAGTACCGCTTCCAGGTGTACGTGAAGCAGCCCGCGGACATCCCG 241
 Db 863 TGTCCCCCAGTTTCAGTACCGCTTCCAGGTGTACGTGAAGCAGCCCGCGGACATCCCG 922
 QY 242 ACTACAAGAAAGTCTGCTTCCCCGAGGGCTTCAAGTGGGAGCGGCTGATGAACCTTCGAGG 301
 Db 923 ACTACAAGAAAGTCTGCTTCCCCGAGGGCTTCAAGTGGGAGCGGCTGATGAACCTTCGAGG 982
 QY 302 ACGGCGGGTGGGACCGTGACCGTGAACCGAGTCTCTCCCTGCGAGGACGGTGTCTTCACTACA 361
 Db 983 ACGGCGGGTGGTGAACCGTGACCGTGAACCGAGTCTCTCCCTGCGAGGACGGTGTCTTCACTACA 1042
 QY 362 AGGTGAAGTTTCATCGGGGTGAACCTTCCCCGAGGGCTTCAAGTGGGAGCGGCTGATGAAGAAAGACCA 421
 Db 1043 AGGTGAAGTTTCATCGGGGTGAACCTTCCCCGAGGGCTTCAAGTGGGAGCGGCTGATGAAGAAAGACCA 1102
 QY 422 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGCGAGCGGCTGCTGAAAGGGCGAGA 481
 Db 1103 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGCGAGCGGCTGCTGAAAGGGCGAGA 1162
 QY 482 TCCACAGGCGCTGAAGCTGAAGGACGGCGGCGCACTACCTGCTGGAGTTCAGTCCATCT 541
 Db 1163 TCCACAGGCGCTGAAGCTGAAGGACGGCGGCGCACTACCTGCTGGAGTTCAGTCCATCT 1222
 QY 542 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTACTACTACTACTACTACTACTACT 601
 Db 1223 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTACTACTACTACTACTACTACTACT 1282
 QY 602 TCACCTCCCAACAGGAGCTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 661
 Db 1283 TCACCTCCCAACAGGAGCTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 1342
 QY 662 ACCACCTGTTCCCTGTA 677
 Db 1343 ACCACCTGTTCCCTGTA 1358

RESULT 10

ACC44640

ID ACC44640 standard; DNA; 4692 BP.

XX AC ACC44640;

XX DT 29-MAY-2003 (first entry)

XX DE Vector pDsRed1N1 nucleotide sequence SEQ ID NO:29.

XX Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
 XX att site; integrase; recombinase; ACes; gene therapy; transgenic animal;
 KW platform artificial chromosome expression system; gene; ds.

XX		
OS	Dicosoma sp.	
OS	Synthetic.	
XX		
PX	WO200297059-A2.	
PX		
PD	05-DEC-2002.	
PD		
PF	30-MAY-2002; 2002WO-US017452.	
XX		
PR	30-MAY-2001; 2001US-0294758P.	
PR	21-MAR-2002; 2002US-0366891P.	
XX	(CHRO-) CHROMOS MOLECULAR SYSTEMS INC.	
PA		
PI	Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;	
PI	Stewart S, Shellard J;	
XX		
DR	WPI; 2003-140461/13.	
XX		
PT	Novel eukaryotic chromosome comprising one or many att sites which	
PT	permits site-directed integration in the presence of lambda-integrase,	
PT	useful for site-specific recombination-directed integration of DNA of	
PT	interest.	
XX		
XX	Example 1; Page 182-184; 272pp; English.	
XX		
CC	The present invention describes a eukaryotic chromosome (I) comprising	
CC	one or several att sites, where an att site is heterologous to the	
CC	chromosome and permits site-directed integration in the presence of	
CC	lambda-integrase. Also described: (1) a platform artificial chromosome	
CC	expression system (Aces) (II) comprising several sites that participate	
CC	in recombinase catalysed recombination; and (2) a method (M1) for	
CC	introducing a heterologous nucleic acid into a platform artificial	
CC	chromosome. (1) can be used in gene therapy. (M1) is useful for	
CC	introducing a heterologous nucleic acid molecule into a platform for	
CC	artificial chromosome, preferably an Aces. (II) is useful for producing a	
CC	transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or	
CC	mammal) by introducing (II) by cell fusion, lipid-mediated transfection	
CC	by a carrier system, microinjection, microcell fusion, electroporation,	
CC	microprojectile bombardment or direct DNA transfer into an embryonic	
CC	cell, preferably a stem cell or an embryo. (II) comprises a heterologous	
CC	nucleic acid that encodes a therapeutic product which is useful for	
CC	making a library of Aces comprising random portions of a genome. ACC44612	
CC	to ACC44732 and ABP96650 to ABP96657 represent sequences used in the	
CC	exemplification of the present invention	
XX		
XX	Sequence 4692 BP; 1111 A; 1268 C; 1247 G; 1066 T; 0 U; 0 Other;	
XX		
Query Match	97.6%; Score 661.6; DB 7; Length 4692;	
Best Local Similarity	98.7%; Pred. No. 1.1e-106;	
Matches	667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;	
QY	2 TGCGCTCCTCGAGAACGTCATCACCGAGTTCATGCCTTCAAGGTGGCGCATGGAGGGCA 61	
Db	683 TC CGCTCTCTCCAAGAAGTCATCAAGGAGTTTCATGCCTTCAAGGTGGCGCATGGAGGGCA 742	
QY	62 CGGTGAACGGGCCACGAGTTTCAGAGTTCGAGGCGCAGGCGCGGCCCTCTACGAGGGCC 121	
Db	743 CGTGTAACGGGCCACGAGTTTCAGATTCGAGGCGAGGCGCGGCCCTCTACGAGGGCC 802	
QY	122 ACAACACCGTGAAGTTGAAGGTGAACAAGGGCGGCCCTCTGCCCTTCGCTGTGGACATCC 181	
Db	803 ACACACACCGTGAAGTGAAGGTGAACAAGGGCGGCCCTCTGCCCTTCGCTGTGGACATCC 862	
QY	182 TGCTCCCCCGAGTTCACGATACGGCTCCAAGGTGTACGTGAAGCACCCCGCGGCATCCCCG 241	
Db	863 TGCTCCCCCGAGTTCACGATACGGCTCCAAGGTGTACGTGAAGCACCCCGCGGCATCCCCG 922	
QY	242 ACTACAAGAAGCTGTCTCTCCCGAGGGCTTCAAAGTGGAGCGCGTGTGAACACTTCGAGG 301	
Db	923 ACTACAAGAAGCTGTCTCTCCCGAGGGCTTCAAAGTGGAGCGCGTGTGAACACTTCGAGG 982	

Qy	302	ACGGCGCGTGGCGA	CCGTGACCGAGACTCTCCCTGCAGGACGGCTGTTCAATCTACA	361
Db	983	ACGGCGCGTGGTGA	CCGTGACCGAGACTCTCCCTGCAGGACGGCTGTTCAATCTACA	1042
Qy	362	AGTGTAAGTTTCAT	TCGGCGTGAACTTCCCTTCCGACGGCCCGTGATCGACAGAGAAGACCA	421
Db	1043	AGTGTAAGTTTCAT	TCGGCGTGAACTTCCCTTCCGACGGCCCGTAATCGAAGAAGACCA	1102
Qy	422	TGGGCTGGGAGGGCCT	CCACCGAGCGCTGTATCCCCCGCGACGGCGTCTGAAAGGCGGAGA	481
Db	1103	TGGGCTGGGAGGGCCT	CCACCGAGCGCTGTATCCCCCGCGAGCGCGTCTGAAAGGCGGAGA	1162
Qy	482	TCACAAGGCCCTGAA	GCTGAAAGACGGCGGCGCACTACCTGGTGGAGTTTCAAGTCCCATCT	541
Db	1163	TCCACAAGGCCCTGAA	GCTGAAAGACGGCGGCGCACTACCTGGTGGAGTTTCAAGTCCCATCT	1222
Qy	542	ACATGCCCAAGAGAC	CCGTGCAGCTCCCGGCTACTACTACGTGGACACCAAGCTGGACA	601
Db	1223	ACATGCCCAAGAGAC	CCGTGCAGCTCCCGGCTACTACTACGTGGAGTCCCAAGCTGGACA	1282
Qy	602	TCACCTCCCAACAC	GAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGCC	661
Db	1283	TCACCTCCCAACAC	GAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGCC	1342
Qy	662	ACCACCTGTTCCCTGTA	677	
Db	1343	ACCACCTGTTCCCTGTA	1358	

chromosome, and permits site-directed integration in the presence of lambda-integrase. Also described: (I) a platform artificial chromosome expression system (Aces) (II) comprising several sites that participate in recombinase catalysed recombination; and (2) a method (M1) for introducing a heterologous nucleic acid into a platform artificial chromosome. (I) can be used in gene therapy. (M1) is useful for introducing a heterologous nucleic acid molecule into a platform artificial chromosome, preferably an Aces. (II) is useful for producing a transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or mammal) by introducing (II) by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer into an embryonic cell, preferably a stem cell or an embryo. (II) comprises a heterologous nucleic acid that encodes a therapeutic product which is useful for making a library of Aces comprising random portions of a genome. ACC44612 to ACC44732 and ABP96650 to ABP96657 represent sequences used in the exemplification of the present invention

Seq	Sequence	4592 BP;	1111 A;	1268 C;	1247 G;	1066 T;	0 U;	0 Other;
	Query Match	97.6%;	Score	661.6;	DB	7;	Length	4692;
	Best Local Similarity	98.7%;	Pred. No.	1.1e-106;				
	Matches	667;	Conservative	0;	Mismatches	9;	Indels	0;
	Gaps	0;						
QY	2	TGGCTCTCTCGAGAACCTCATCCGAGTTTCATTCGCTTTCAAGGTGCGCATGGAGGGCA	61					
Db	683	TGCGCTCTCTCCAAGAACGTCATCAAGGAGTTTCATTCGCTTCAAGGTGCGCATGGAGGGCA	742					
QY	62	CGGTGAACGGCCACGAGTTTCAGATTCGAGGGCGAGGGCGCGGCCGCCCTACGAGGGCC	121					
Db	743	CGGTGAACGGCCACGAGTTTCAGATTCGAGGGCGAGGGCGCGGCCGCCCTACGAGGGCC	802					
QY	122	ACAAACACCGTGAAGTTGAAGGTGACCAAGGCGCGGCCCTGCCCTTCGCTCGGCATCC	181					
Db	803	ACAAACACCGTGAAGTTGAAGGTGACCAAGGCGCGGCCCTGCCCTTCGCTCGGCATCC	862					
QY	182	TGTCCCCCAGTTCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCGCATCCCGC	241					
Db	863	TGTCCCCCAGTTCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCGCATCCCGC	922					
QY	242	ACTACAAGAGCTGTCCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGTGAACCTTCGAGG	301					
Db	923	ACTACAAGAGCTGTCCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGTGAACCTTCGAGG	982					

control over fertility and reproduction. The construct comprises a native promoter, a blocking DNA sequence contoured for and designed to abrogate a crucial gene's function or to cause its mis-expression, and a genetic switch to regulate controlled expression/repression of the blocker/gene knockout. The construct is useful for preventing embryogenesis or gametogenesis in animals by stably transforming an animal cell with the construct by microinjection, transfection or infection, where the construct stably integrates into the genome by homologous recombination, and implanting the cell into a host organism, where a whole animal develops from the implanted cell. The present DNA sequence is plasmid construct pBir(DHSP)-RFP-ohoxDS/SH used for transfection of Pacific oysters. The plasmid comprises oyster Hox gene under the under the control of Drosophila heat shock promoter (dhSP), red fluorescent protein and SV40 PolyA

XX SQ Sequence 7910 BP; 2016 A; 1893 C; 1924 G; 2077 T; 0 U; 0 Other;

Query Match 97.6%; Score 661.6; DB 4; Length 7910;
Best Local Similarity 98.7%; Pred. No. 1.1e-106;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

2 TGGCTCTCTCGAGAACCTCATCCAGGTTTCATGGCGTTTCAAGTGCGCATGGAGGCA 61
2704 TGGCTCTCTCGAGAACCTCATCCAGGTTTCATGGCGTTTCAAGTGCGCATGGAGGCA 2645

62 CCGTGAACCGGCACAGTTTCAGATCGAGGCGAGGCGGAGGCGCCCGCCCTACGAGGGCC 121
2644 CCGTGAACCGGCACAGTTTCAGATCGAGGCGAGGCGGAGGCGCCCGCCCTACGAGGGCC 2585

122 ACACACCGTGAAGTTCAGAGTGACCAAGGCGGCGCCCTGCTTCCGCTGGGACATCC 181
2584 ACACACCGTGAAGTTCAGAGTGACCAAGGCGGCGCCCTGCTTCCGCTGGGACATCC 2525

182 TGTCCCCCGAGTTCCAGTCCAGTTCAGGTGACGAGGAGACACCCCGCGGACATCCCCG 241
2524 TGTCCCCCGAGTTCCAGTTCAGGTGACGAGGAGACACCCCGCGGACATCCCCG 2465

242 ACTACAGAGAGCTGCTCTCCCGAGGCTTCAAGTGGGAGCGGTGATGAATCTTCAGG 301
2464 ACTACAGAGAGCTGCTCTCCCGAGGCTTCAAGTGGGAGCGGTGATGAATCTTCAGG 2405

302 ACGGCGGCGTGGCGACCGTGAACCCAGGACTCTCCCTGCGAGGAGCGGTGCTTCATCTACA 361
2404 ACGGCGGCGTGGTGAACCGTGAACCCAGGACTCTCCCTGCGAGGAGCGGTGCTTCATCTACA 2345

362 AGGTGAAGTTTCATCGGCGTGAATCTCCCTCGAGCGGCCCGGTGATGACAGAGACCA 421
2344 AGGTGAAGTTTCATCGGCGTGAATCTCCCTCGAGCGGCCCGGTGATGACAGAGACCA 2285

422 TGGGCTGGGAGGCTTCCACCGAGGCGCTGTACCCCGCGGCGGTGCTGAAGGCGGAGA 481
2284 TGGGCTGGGAGGCTTCCACCGAGGCGCTGTACCCCGCGGCGGTGCTGAAGGCGGAGA 2225

482 TCCACAGAGGCGCTGAAGTGAAGAGCGGCGGACCTACCTGGTGGAGTTCAAGTCCATCT 541
2224 TCCACAGAGGCGCTGAAGTGAAGAGCGGCGGACCTACCTGGTGGAGTTCAAGTCCATCT 2165

542 ACATGGCCAGAGCGGCTGCGAGTCCCGGCTACTACTGAGGACCAACCAAGCTGACA 601
2164 ACATGGCCAGAGCGGCTGCGAGTCCCGGCTACTACTGAGGACCTCAAGCTGACA 2105

602 TCACCTCCCAACAGGAGGACTTACACCTCGTGGAGCAGTACGAGCGCACCGAGGCGGCC 661
2104 TCACCTCCCAACAGGAGGACTTACACCTCGTGGAGCAGTACGAGCGCACCGAGGCGGCC 2045

662 ACCACCTGTTCTCTGA 677
2044 ACCACCTGTTCTCTGA 2029

RESULT 12
ABS56664
ID ABS56664 standard; DNA; 9320 BP.

XX AC ABS56664;
XX DT 24-JAN-2003 (first entry)
XX DE Plasmid p5-Puro-CMV-(N'-EGFP)-CMV-Red(EGFP-EJ) DNA.
XX KW Vector; genotoxicity; food component; cosmetic; medicine; cancer;
XX KW Genetic stability; hygromycin; ds.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT misc_feature 1..1592
XX FT /tag= a
XX FT /note= "retroviral vector p5NM"
XX FT CDS 1617..2216
XX FT /tag= b
XX FT /product= "puromycin resistance gene from pRetroOn"
XX FT promoter 2267..2848
XX FT /tag= c
XX FT /note= "CMV promoter from pEGFP-N1"
XX FT CDS 2906..3348
XX FT /tag= d
XX FT /product= "N'-EGFP"
XX FT misc_feature 3374..3392
XX FT /tag= e
XX FT /note= "retroviral vector p5NM"
XX FT promoter 3411..3992
XX FT /tag= f
XX FT /note= "CMV promoter from pDsRed1-N1"
XX FT CDS 4038..4718
XX FT /tag= g
XX FT /product= "pDsRed1-N1"
XX FT CDS 4766..5508
XX FT /tag= h
XX FT /product= "EGFP-EJ"
XX FT misc_feature 5527..9320
XX FT /tag= i
XX FT /note= "retroviral vector p5NM"
XX PN WO200270740-A2.
XX PD 12-SEP-2002.
XX PF 28-FEB-2002; 2002WO-EP002194.
XX PR 05-MAR-2001; 2001DE-01010449.
XX (WIES/) WIESMUELLER L.
XX Wiesmuelle L;
XX WPI; 2002-682909/73.
XX New vector containing recombination sequences, useful e.g. for assessing compounds for genotoxicity, cancer susceptibility and mutational frequency.
XX Disclosure; Page 104-106; 106pp; German.
XX This invention describes a novel vector containing, at least, two homologous DNA sequences that, through DNA exchange, induce a change in a gene present in the vector. The products of the invention are used to assess genotoxicity of compounds (e.g. components of foods, cosmetics and medicines), to determine susceptibility for development or progression of cancer, for analysis of (non-)conservative recombination processes, and joining and inactivating/reverting mutations, for characterization of selected genes with respect to maintenance or reduction of genetic stability, and to determine genetic (in)stability and recombinational or mutational frequency of cells, tissues or organisms. The vector allows an analysis to be performed in hours (making it suitable for large-scale, routine use), and can be used with a variety of eukaryotic cells and

CC animals. All known types of DNA recombinations can be examined. This
 CC sequence represents a plasmid p5-Puro-CMV-(N'-EGFP)-CMV-Red(EGFP-EU) DNA,
 CC described in the disclosure of the invention
 XX
 SQ Sequence 9320 BP; 2124 A; 2687 C; 2474 G; 2035 T; 0 U; 0 Other;

Query Match 97.6%; Score 661.6; DB 6; Length 9320;
 Best Local Similarity 98.7%; Pred. No. 1.1e-106;
 Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGCCTCCCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGCA 61
 DB |||||
 4042 TGGCTCTCTCCAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGCA 4101
 QY 62 CCGTGAACCGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCGCCCTACAGAGGCC 121
 DB |||||
 4102 CCGTGAACCGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCGCCCTACAGAGGCC 4161
 QY 122 ACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGCCCTCGCTTCCGCTGGGACATCC 181
 DB |||||
 4162 ACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGCCCTCGCTTCCGCTGGGACATCC 4221
 QY 182 TGTCCCCCAGTTCAGTACGGCTTCAAGTGTAGTGAAGCACCCCGCGACATCCCG 241
 DB |||||
 4222 TGTCCCCCAGTTCAGTACGGCTTCAAGTGTAGTGAAGCACCCCGCGACATCCCG 4281
 QY 242 ACTACAGAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATTCGAGG 301
 DB |||||
 4282 ACTACAGAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATTCGAGG 4341
 QY 302 ACGGGCGGTGGCGACCGTGAACCCAGGACTCTCTCCCTGAGGACGGCTGCTTCATCTACA 361
 DB |||||
 4342 ACGGGCGGTGGCGACCGTGAACCCAGGACTCTCTCCCTGAGGACGGCTGCTTCATCTACA 4401
 QY 362 AGGTGAAGTTTCATCGCGTGAATCTCCCTCCGAGGGCGCCCTGATGAGAGAGACA 421
 DB |||||
 4402 AGGTGAAGTTTCATCGCGTGAATCTCCCTCCGAGGGCGCCCTGATGAGAGAGACA 4461
 QY 422 TGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGAGCGCTGTGAAGGGCGAGA 481
 DB |||||
 4462 TGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGAGCGCTGTGAAGGGCGAGA 4521
 QY 482 TCCACAAGCCCTGAAGCTGAAGGACGGCGGCACCTACTGTTGGAGTTCAAGTCCATCT 541
 DB |||||
 4522 TCCACAAGCCCTGAAGCTGAAGGACGGCGGCACCTACTGTTGGAGTTCAAGTCCATCT 4581
 QY 542 ACATGGCCAGAGACCGCTGAGCTGCGCGGTACTACTACTGACACACCAAGCTGGACA 601
 DB |||||
 4582 ACATGGCCAGAGACCGCTGAGCTGCGCGGTACTACTACTGACACACCAAGCTGGACA 4641
 QY 602 TCACCTCCCAACAGGAGTACACCATCGTGGAGCAGTACAGGCGCACCGAGGCGCGCC 661
 DB |||||
 4642 TCACCTCCCAACAGGAGTACACCATCGTGGAGCAGTACAGGCGCACCGAGGCGCGCC 4701
 QY 662 ACCACTGTTCCTGTA 677
 DB |||||
 4702 ACCACTGTTCCTGTA 4717

RESULT 13

AAD10003

ID AAD10003 standard; DNA; 5436 BP.

XX

AC AAD10003;

XX

DT 12-SEP-2001 (first entry)

XX

DE Plasmid pSPW6 to express goosecoid cDNA fused to red fluorescent protein.

XX

KW Mouse; fertility; reproduction; gametogenesis; microinjection; CMV;

KW cytomegalovirus; SV40; simian virus 40; infection; goosecoid gene;

KW plasmid pSPW 6; embryogenesis; ds.

XX

OS Mus musculus.
 OS Cytomegalovirus.
 OS Rhesus macaque polyoma virus.
 OS Unidentified.
 OS Chimeric.

XX WO200148224-A1.

PN 05-JUL-2001.

XX 22-DEC-2000; 2000WO-AU001596.

XX 24-DEC-1999; 99AU-00004884.

PR (CSIR) COMMONWEALTH SCI & IND RES ORG.

XX Thresher R, Hinds L, Hardy C, Whyard S, Vignarajan S, Grewe PM;

PI Patil J;

XX WPI; 2001-425672/45.

XX Novel construct for preventing embryogenesis in animals comprises native

PT promoter, blocking DNA which abrogates function of crucial gene and

PT Genetic switch to regulate expression/repression of blocker/gene

PT knockout.

XX Example 13; Page 215-217; 241pp; English.

PS The invention relates to a construct which allows animals to be bred in

XX captivity but renders them infertile in the wild by allowing reversible

CC control over fertility and reproduction. The construct comprises a native

CC promoter, a blocking DNA sequence contoured for and designed to abrogate

CC a crucial gene's function or to cause its mis-expression, and a genetic

CC switch to regulate controlled expression/repression of the blocker/gene

CC knockout.

CC The construct is useful for preventing embryogenesis or

CC gametogenesis in animals by stably transforming an animal cell with the

CC construct by microinjection, transfection or infection, where the

CC construct stably integrates into the genome by homologous recombination,

CC and implanting the cell into a host organism, where a whole animal

CC develops from the implanted cell. The present sequence is plasmid pSPW 6

CC containing mouse goosecoid promoter expressing goosecoid cDNA fused to

CC red fluorescent protein. The plasmid contains CMV promoter and SV40 poly

CC and SV40 ori

XX Sequence 5436 BP; 1247 A; 1525 C; 1492 G; 1172 T; 0 U; 0 Other;

SQ

Query Match 97.3%; Score 660; DB 4; Length 5436;

Best Local Similarity 98.5%; Pred. No. 2.2e-106;

Matches 666; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 TGGCCTCTCCGAGAACGTCATCACCGAGTTTCATGGCGCTTCAAGTGGCGATGGAGGCA 61

DB |||||

1427 TGGCTCTCTCCAGAACGTCATCAAGAGTTTCATGGCTTCAAGTGGCGATGGAGGCA 1486

QY 62 CCGTGAACCGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCGCCCTACAGAGGCC 121

DB |||||

1487 CCGTGAACCGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCGCCCTACAGAGGCC 1546

QY 122 ACAACACCGTGAAGTTTCGAGTGAACCAAGGGCGGCCCTCGCTTCCGCTGGGACATCC 181

DB |||||

1547 ACAACACCGTGAAGTTTCGAGTGAACCAAGGGCGGCCCTCGCTTCCGCTGGGACATCC 1606

QY 182 TGTCCCCCAGTTCAGTACGGCTTCAAGTGTAGTGAAGCACCCCGCGACATCCCG 241

DB |||||

1607 TGTCCCCCAGTTCAGTACGGCTTCAAGTGTAGTGAAGCACCCCGCGACATCCCG 1666

QY 242 ACTACAGAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATCAATTCGAGG 301

DB |||||

1667 ACTACAGAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATCAATTCGAGG 1726

QY 302 ACGGCGCGGTGGGACCGCTGACCCAGGACTCTCCCTCGCAGGAGCGGTGCTTCATCTACA 361

DB |||||

1727 ACGGCGCGGTGGTGAACCGTGAACCCAGGACTCTCCCTCGCAGGAGCGGTGCTTCATCTACA 1786

```

QY 362 AGGTGAAGTTTCATCGGCGTAACTTCCCTCCGACGGCCCGTGGATGCAGAGAACCA 421
Db 1787 AGGTGAAGTTTCATCGGCGTAACTTCCCTCCGACGGCCCGTAAATGCAGAGAACCA 1846
QY 422 TGGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTGTGAAGGGCGACA 481
Db 1847 TGGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTGTGAAGGGCGACA 1906
QY 482 TCACAAAGGCGCTCAAGCTGAAGAGCGCGGCCACTACTCTGTGGAGTTCAAGTCCATCT 541
Db 1907 TCACAAAGGCGCTCAAGCTGAAGAGCGCGGCCACTACTCTGTGGAGTTCAAGTCCATCT 1966
QY 542 ACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACACCAAGCTGGACA 601
Db 1967 ACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACA 2026
QY 602 TCACCTCCCAACAGAGGACTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGCC 661
Db 2027 TCACCTCCCAACAGAGGACTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGCC 2086
QY 662 ACCACCTGTTCTGTGA 677
Db 2087 ACCACCTGTTCTGTGA 2102

RESULT 14
ID ADE24111 standard; DNA; 6893 BP.
XX AC ADE24111;
XX DT 29-JAN-2004 (first entry)
XX DE Proviral plasmid DNA SFR.
XX KW site-specific DNA recombination; Cre recombinase;
XX KW protein-transduction domain; gene activation; gene inactivation;
XX KW chromosomal translocation; reporter protein; ds; circular.
XX OS Synthetic.
XX FH Key
XX FT misc_RNA
XX FT repeat_region
XX FT primer_bind
XX FT 5'UTR
XX FT misc_RNA
XX FT CDS
XX FT misc_RNA
XX FT CDS
XX FT misc_RNA
XX FT CDS
XX FT CDS
XX FT 3'UTR

```

```

FT repeat_region /*tag= j
FT 4081..4641
FT /*tag= k
FT /standard_name= "LTR"
FT /note= "3'-LTR"
FT 4642..6893
FT /*tag= l
FT /note= "plasmid pUC fragment"
XX WO2003070931-A2.
XX 28-AUG-2003.
XX 19-FEB-2003; 2003WO-EP001680.
XX 21-FEB-2002; 2002DE-01007313.
XX 16-JUL-2002; 2002DE-01032196.
XX (VISI-) VISION 7 GMBH.
XX Baum C, Will E, Ostertag W, Klump H, Schiedlmeier B;
XX WPI; 2003-767353/72.
XX In vitro or in vivo site-specific DNA recombination, useful e.g. for gene
XX inactivation, using Cre recombinase that lacks heterologous protein-
XX transduction domain.
XX Claim 23; SEQ ID NO 8; 84pp; German.
XX This invention describes a novel method of site-specific DNA
XX recombination in eukaryotic cells in vitro comprising using a
XX bacteriophage P1 Cre recombinase that lacks heterologous protein-
XX transduction domains. The invention also contains a reporter system for
XX detecting site-specific DNA recombination in eukaryotic cells.
XX Recombination with Cre recombinase lacking heterologous protein-
XX transduction domains is used for therapeutic site-specific recombination
XX in eukaryotic cells, in vivo or in vitro, e.g. for partial inactivation
XX or activation of genes, also to create a chromosomal translocation. The
XX modified recombinase can enter intact (not electrically or chemically
XX treated) cells when added to culture medium and catalyzes recombination
XX events in the nucleus with recombination rate over 50%, after a single
XX application. The Cre-recombinase does not need to be supplied from
XX nucleic acid introduced into the cell, so its effect is immediate but
XX short-lived, limiting cytotoxic effects; eliminating the risk of non-
XX specific integration of DNA into the genome and providing quick results.
XX This sequence represents the proviral plasmid Cre reporter construct SFR
XX used in the method described in the invention.
XX Sequence 6893 BP; 1564 A; 1940 C; 1778 G; 1611 T; 0 U; 0 Other;
XX
Query Match 97.3%; Score 659.6; DB 9; Length 6893;
Best Local Similarity 98.7%; Pred. No. 2.5e-106;
Matches 665; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGCTCTCTCCGAGAACGTCTATCCCGAGTTTCATCGGCTTCAAGTGGCGCATGGAGGCA 61
Db 1604 TCGGCTCTCTCCGAGAACGTCTATCCCGAGTTTCATCGGCTTCAAGTGGCGCATGGAGGCA 1663
QY 62 CCGTGAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCCTTACGAGGGCC 121
Db 1664 CCGTGAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCCTTACGAGGGCC 1723
QY 122 ACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGCCCTTCCCTTGGGACATCC 181
Db 1724 ACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGCCCTTCCCTTGGGACATCC 1783
QY 182 TGTCCCCCAGTTCAGTACCGCTTCCAGGTGTACGTGAGACCCCGGACATCCCG 241
Db 1784 TGTCCCCCAGTTCAGTACCGCTTCCAGGTGTACGTGAGACCCCGGACATCCCG 1843
QY 242 ACTACAAGAAGCTGTCTTCCCCCGAGGGCTTCAAGTGGGCGCGTGTATGAATTCGAGG 301

```


Db 605 TCACCTCCCAACAGGAGTACACCATCGTGGAGCAGTACGAGCGCAGCGCGCC 664
QY 662 ACCACCTGTTCTTGTA 677
Db 665 ACCACCTGTTCTTGGA 680

Search completed: August 3, 2004, 11:29:37
Job time : 453.858 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: August 3, 2004, 10:03:29 ; Search time 4535.5 Seconds
(without alignments)
6479.234 Million cell updates/sec
Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggctctctccgagaagct.....gccaccactgttctctgttaa 678
Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : GenEmbl:
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	678	100.0	678	6	AX824732 Sequence
2	676.4	99.8	678	6	AX370408 Sequence
3	672.2	99.1	1050	6	AX666133 Sequence
4	670.2	98.8	675	6	AX824731 Sequence
5	668.4	98.6	678	6	AX370406 Sequence
6	665.2	98.1	678	6	AX370404 Sequence
7	665.2	98.1	678	6	AX824725 Sequence
8	661.6	97.6	4692	6	AX463702 Sequence
9	661.6	97.6	9320	6	AX663075 Sequence
10	659.6	97.3	6893	6	AX823860 Sequence
11	649.8	95.8	10141	12	AY342347 Red H-Pel
12	649.8	95.8	10276	12	AY342348 Red H-Pel
13	629.6	92.9	681	12	AF506025 Synthetic
14	627.6	92.6	1395	12	AF506026 Synthetic
15	614.2	90.6	681	6	AX233584 Sequence
16	614.2	90.6	713	6	AX233627 Sequence
17	568.6	83.9	678	12	AF506027 Synthetic
18	556.4	82.1	678	6	AX824729 Sequence
19	438	64.6	678	6	AR183915 Sequence
20	438	64.6	678	6	AX172854 Sequence
21	438	64.6	678	6	AX207715 Sequence
22	438	64.6	678	6	AX233581 Sequence
23	438	64.6	859	3	AF168419 Discosoma
24	438	64.6	859	6	AX463698 Sequence
25	432	63.7	666	6	AX348043 Sequence
26	432	63.7	666	6	AX353910 Sequence
27	422	62.2	898	6	AX686888 Sequence
28	377.8	55.7	876	3	AF272711 Discosoma
29	377.8	55.7	876	6	AX686894 Sequence
30	313.4	46.2	707	6	AX824739 Sequence
31	308.6	45.5	699	12	AY233273 Synthetic
32	305	45.0	4707	12	AY218848 Monster G
33	297.2	43.8	654	6	AX824741 Sequence
34	296.6	43.7	669	6	AX699925 Sequence
35	295	43.5	669	6	AX699921 Sequence
36	295	43.5	669	6	AX699927 Sequence
37	293.4	43.3	660	6	AX699761 Sequence
38	293.4	43.3	693	6	AX699793 Sequence
39	292.4	43.1	660	6	AX699825 Sequence
40	292.4	43.1	687	6	AX527900 Sequence
41	292.4	43.1	1396	6	AX527902 Sequence
42	292.4	43.1	1424	6	AX527904 Sequence
43	291.8	43.0	660	6	AX699785 Sequence
44	291.8	43.0	660	6	AX699809 Sequence
45	291.8	43.0	663	6	AX699755 Sequence

ALIGNMENTS

RESULT 1
AX824732
LOCUS AX824732 678 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 14 from Patent WO02068459.
ACCESSION AX824732
VERSION AX824732.1 GI:39750595
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Non aggregating fluorescent proteins and methods for using the same
Patent: WO 02068459-A 14 06-SEP-2002;
Location/Qualifiers

SOURCE	synthetic construct
ORGANISM	synthetic construct
TITLE	artificial sequences.
JOURNAL	
FEATURES	1 Non aggregating fluorescent proteins and methods for using the same Patent: WO 02068459-A 13 06-SEP-2002;
source	Location/Qualifiers 1..675 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="non-aggregating mutant"
ORIGIN	
Query Match	98.8%; Score 670.2; DB 6; Length 675;
Best Local Similarity	99.6%; Pred. No. 5.8e-79;
Matches 672; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	1 ATGGGCTCCTCCGAGAACGTTCATCACCGAGTTTCAAGGTCGCATCGAGGCC 60
DB	1 ATGGGCTCCTCCGAGAACGTTCATCACCGAGTTTCAAGGTCGCATCGAGGCC 60
QY	61 ACCGTGAACGGCCACGAGTTTCAGATTCGAGGCGGAGGGCGGCCCTACGAGGCC 120
DB	61 ACCGTGAACGGCCACGAGTTTCAGATTCGAGGCGGAGGGCGGCCCTACGAGGCC 120
QY	121 CACAACACCGTGAAAGTGAAGGTGACCAAGGCGGCGCCCCCTTGCCTGGGACATC 180
DB	121 CACAACACCGTGAAAGTGAAGGTGACCAAGGCGGCGCCCCCTTGCCTGGGACATC 180
QY	181 CTGTCCCCCAGTTTCAGTAACGCTTCAAAGGTGTAAGTAGAAGCACCAGCGCATCCCC 240
DB	181 CTGTCCCCCAGTTTCAGTAACGCTTCAAAGGTGTAAGTAGAAGCACCAGCGCATCCCC 240
QY	241 GACTACAAGAAGCTGTCTTCCCAGGAGGCTTCAAAGTGGGAGCGGTGATCAACTTCGAG 300
DB	241 GACTACAAGAAGCTGTCTTCCCAGGAGGCTTCAAAGTGGGAGCGGTGATCAACTTCGAG 300
QY	301 GACGGCGGGTGGCGACCGTGACCCAGAGCTCTCTCCCTGCAGGACGGCTTCACTTAC 360
DB	301 GACGGCGGGTGGCGACCGTGACCCAGAGCTCTCTCCCTGCAGGACGGCTTCACTTAC 360
QY	361 AAGTGAAGTTTCACTGGGCTGAACCTTCCCTCCGAGCGCCCGGTATCAGAGAGACC 420
DB	361 AAGTGAAGTTTCACTGGGCTGAACCTTCCCTCCGAGCGCCCGGTATCAGAGAGACC 420
QY	421 ATGGGCTGGAGGCTTCCACCGAGCGCTGTACCCCGCGACGGCGTCTGAAAGGGCGAG 480
DB	421 ATGGGCTGGAGGCTTCCACCGAGCGCTGTACCCCGCGACGGCGTCTGAAAGGGCGAG 480
QY	481 ATCCAACAGGCGCTGAAGCTGAAGGAGCGCGGCGCACTACTGTGTGGATTCAAGTCCATC 540
DB	481 ACCCAACAGGCGCTGAAGCTGAAGGAGCGCGGCGCACTACTGTGTGGATTCAAGTCCATC 540
QY	541 TACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
DB	541 TACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
QY	601 ATCACTTCCCAACAGGAGCTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
DB	601 ATCACTTCCCAACAGGAGCTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
QY	661 CACCACCTGTTCTCTG 675
DB	661 CACCACCTGTTCTCTG 675
RESULT 5	
AX370406	
LOCUS	AX370406 678 bp DNA linear PAT 16-FEB-2002
DEFINITION	Sequence 3 from Patent W00196373.
ACCESSION	AX370406
KEYWORDS	
REFERENCE	1 Zhao, M., Xu, M., Jiang, P. and Yang, M. Fluorescent proteins Patent: WO 02060941-A 7 08-AUG-2002; Anticancer, Inc. (US)
FEATURES	Location/Qualifiers 1..1050 /organism="unidentified" /mol_type="unassigned DNA" /db_xref="taxon:32644" /note="Coral"
ORIGIN	
Query Match	99.1%; Score 672.2; DB 6; Length 1050;
Best Local Similarity	99.6%; Pred. No. 2.9e-79;
Matches 674; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	1 ATGGGCTCCTCCGAGAACGTTCATCACCGAGTTTCAAGGTCGCATCGAGGCC 60
DB	289 ATGGGCTCCTCCGAGAACGTTCATCACCGAGTTTCAAGGTCGCATCGAGGCC 348
QY	61 ACCGTGAACGGCCACGAGTTTCAGATTCGAGGCGGAGGGCGGCCCTACGAGGCC 120
DB	349 ACCGTGAACGGCCACGAGTTTCAGATTCGAGGCGGAGGGCGGCCCTACGAGGCC 408
QY	121 CACAACACCGTGAAAGTGAAGGTGACCAAGGCGGCGCCCCCTTGCCTGGGACATC 180
DB	409 CACAACACCGTGAAAGTGAAGGTGACCAAGGCGGCGCCCCCTTGCCTGGGACATC 468
QY	181 CTGTCCCCCAGTTTCAGTAACGCTTCAAAGGTGTAAGTAGAAGCACCAGCGCATCCCC 240
DB	469 CTGTCCCCCAGTTTCAGTAACGCTTCAAAGGTGTAAGTAGAAGCACCAGCGCATCCCC 528
QY	241 GACTACAAGAAGCTGTCTTCCCAGGAGGCTTCAAAGTGGGAGCGGTGATCAACTTCGAG 300
DB	529 GACTACAAGAAGCTGTCTTCCCAGGAGGCTTCAAAGTGGGAGCGGTGATCAACTTCGAG 588
QY	301 GACGGCGGGTGGCGACCGTGACCCAGAGCTCTCTCCCTGCAGGACGGCTTCACTTAC 360
DB	589 GACGGCGGGTGGCGACCGTGACCCAGAGCTCTCTCCCTGCAGGACGGCTTCACTTAC 648
QY	361 AAGTGAAGTTTCACTGGGCTGAACCTTCCCTCCGAGCGCCCGGTATCAGAGAGACC 420
DB	649 AAGTGAAGTTTCACTGGGCTGAACCTTCCCTCCGAGCGCCCGGTATCAGAGAGACC 708
QY	421 ATGGGCTGGAGGCTTCCACCGAGCGCTGTACCCCGCGACGGCGTCTGAAAGGGCGAG 480
DB	709 ATGGGCTGGAGGCTTCCACCGAGCGCTGTACCCCGCGACGGCGTCTGAAAGGGCGAG 768
QY	481 ATCCAACAGGCGCTGAAGCTGAAGGAGCGCGGCGCACTACTGTGTGGATTCAAGTCCATC 540
DB	769 ACCCAACAGGCGCTGAAGCTGAAGGAGCGCGGCGCACTACTGTGTGGATTCAAGTCCATC 828
QY	541 TACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
DB	829 TACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 888
QY	601 ATCACTTCCCAACAGGAGCTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
DB	889 ATCACTTCCCAACAGGAGCTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGGC 948
QY	661 CACCACCTGTTCTCTGTA 677
DB	949 CACCACCTGTTCTCTGTA 965
RESULT 4	
AX824731	
LOCUS	AX824731 675 bp DNA linear PAT 11-DEC-2003
DEFINITION	Sequence 13 from Patent W002068459.
ACCESSION	AX824731
VERSION	AX824731.1 GI:39750594
KEYWORDS	

RESULT 4			
AX824731	AX824731	675 bp	DNA
LOCUS	Sequence 13 from Patent WO02068459.		
DEFINITION	AX824731		
ACCESSION	AX824731.1		
VERSION	GI:39750594		
KEYWORDS			

RESULT 5			
AX370406	AX370406	678 bp	DNA
LOCUS	Sequence 3 from Patent WO0196373.		
DEFINITION	AX370406		
ACCESSION	AX370406		

•


```
RESULT 7
AX824725          678 bp      DNA      linear      PAT 11-DEC-2003
LOCUS
DEFINITION
Sequence 7 from Patent WO02068459.
AX824725
ACCESSION
AX824725.1      GI:39750591
VERSION
KEYWORDS
Discosoma sp.
ORGANISM
Discosoma sp.
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE
1
Non aggregating fluorescent proteins and methods for using the same
Patent: WO 02068459-A 7 06-SEP-2002;
JOURNAL
Location/Qualifiers
1. .678
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"
FEATURES
source
ORIGIN
Query Match          98.1%; Score 665.2; DB 6; Length 678;
Best Local Similarity 98.8%; Pred. No. 2.6e-78;
Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGGCTCTCCAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC 60
Db 1 ATGGCTCTCCAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC 60
QY 61 ACCGTGAACGGCCACAGTTCGAGATTCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGC 120
Db 61 ACCGTGAACGGCCACAGTTCGAGATTCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGC 120
QY 121 CACAAACACCGTGAAGTTGAAGTTGACCAAGGGCGGGCCCGCTTGCCTTGGCTGGGACATC 180
Db 121 CACAAACACCGTGAAGTTGAAGTTGACCAAGGGCGGGCCCGCTTGCCTTGGGACATC 180
QY 181 CTGTCCCGCCAGTTCAGTAGCGCTCCAAAGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 CTGTCCCGCCAGTTCAGTAGCGCTCCAAAGTGTACGTGAAGCACCCCGCGACATCCCC 240
QY 241 GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 300
Db 241 GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 300
QY 301 GACGGCGGCTGGGACCGTGACCCAGGACTCCTCCCTGAGAGCGGCTGCTTCATCTAC 360
Db 301 GACGGCGGCTGGGACCGTGACCCAGGACTCCTCCCTGAGAGCGGCTGCTTCATCTAC 360
QY 361 AAGTGAAGTTTCATCGGCGTGAACCTTCCCTCCGACGGCCCGCTGATGCAGAAAGAGCC 420
Db 361 AAGTGAAGTTTCATCGGCGTGAACCTTCCCTCCGACGGCCCGCTGATGCAGAAAGAGCC 420
QY 421 ATGGGTGGAGGCGCTTCAACGAGCGCTGTACCCCGCGACGGCGGTGATGAAGGGGAG 480
Db 421 ATGGGTGGAGGCGCTTCAACGAGCGCTGTACCCCGCGACGGCGGTGATGAAGGGGAG 480
QY 481 ATCCACAAGGCCCTGAAGCTGAGAGCGGGCGGCACTTACCTGGTGGAGTTCAAGTCCATC 540
Db 481 ATCCACAAGGCCCTGAAGCTGAGAGCGGGCGGCACTTACCTGGTGGAGTTCAAGTCCATC 540
QY 541 TACATGGCCAGAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACACCAAGCTGGAC 600
Db 541 TACATGGCCAGAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACACCAAGCTGGAC 600
QY 601 ATCACTCCCAACAGGAGTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
Db 601 ATCACTCCCAACAGGAGTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
QY 661 CACCACTGTTCCTGTAA 678
|||||
```

```
Db 661 CACCACCTGTTCCTGTAA 678
|||||
RESULT 8
AX463702          4692 bp      DNA      linear      PAT 15-JUL-2002
LOCUS
DEFINITION
Sequence 16 from Patent WO0248338.
AX463702
ACCESSION
AX463702.1      GI:21886461
VERSION
KEYWORDS
synthetic construct
synthetic construct
artificial sequences.
ORGANISM
Lichtenberg-Frat,H.
Yeast strain for testing the geno- and cytotoxicity of complex
environmental contamination
Patent: WO 0248338-A 16 20-JUN-2002;
JOURNAL
Lichtenberg-Frat, Hella (DE)
Location/Qualifiers
1. .4692
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/Note="Vektor pDSRed1-N1"
FEATURES
source
ORIGIN
Query Match          97.6%; Score 661.6; DB 6; Length 4692;
Best Local Similarity 98.7%; Pred. No. 5.1e-78;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGCTCTCTCGAGAACGTGATCATCCAGGTTTCATCGCTTCAAGTGGCGATGGAGGCA 61
Db 683 TGGCTCTCTCGAGAACGTGATCATCCAGGTTTCATCGCTTCAAGTGGCGATGGAGGCA 742
QY 62 CGGTGAACGGCCACGAGTTCGAGATTCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGCCC 121
Db 743 CGGTGAACGGCCACGAGTTCGAGATTCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGCCC 802
QY 122 ACAACACCGTGAAGTTGAAGTTGACCAAGGGCGGCGCCCTTGCCTTGGGACATCC 181
Db 803 ACAACACCGTGAAGTTGAAGTTGACCAAGGGCGGCGCCCTTGCCTTGGGACATCC 862
QY 182 TGTCCCGCCAGTTCAGTACCGCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 241
Db 863 TGTCCCGCCAGTTCAGTACCGCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 922
QY 242 ACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 301
Db 923 ACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 982
QY 302 ACGCGGCGTGGCGACCGTGACCCAGGACTCCTCCCTGAGAGCGGCTTCACTTACA 361
Db 983 ACGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGAGAGCGGCTTCACTTACA 1042
QY 362 AGGTGAAGTTTCATCGGCGTGAACCTTCCCGAGGGCCCGCTGATGCAGAAAGAGCA 421
Db 1043 AGGTGAAGTTTCATCGGCGTGAACCTTCCCGAGGGCCCGCTGATGCAGAAAGAGCA 1102
QY 422 TGGGTGGAGGCGCTTCCACCGAGCGCTGTACCCCGCGAGCGGCTGTGAAGGGCGAGA 481
Db 1103 TGGGTGGAGGCGCTTCCACCGAGCGCTGTACCCCGCGAGCGGCTGTGAAGGGCGAGA 1162
QY 482 TCCACAAGGCCCTGAAGCTGAAGGACGGGGCACTTACCTGGTGGAGTTCAAGTCCATCT 541
Db 1163 TCCACAAGGCCCTGAAGCTGAAGGACGGGGCACTTACCTGGTGGAGTTCAAGTCCATCT 1222
QY 542 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACACCAAGCTGGACA 601
Db 1223 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACTCCAAGCTGGACA 1282
QY 602 TACCTTCCCAACAGGAGTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGCC 661
|||||
```

Db 1283 TCACCTCCACACAGGAGCTACACCATCGTGGAGCAGTACAGGCGACCGAGGGCGGCC 1342

QY 662 ACCACCTGTTCTGTGA 677
 |||||
 Db 1343 ACCACCTGTTCTGTGA 1358

RESULT 9
 AX663075
 LOCUS AX663075
 DEFINITION Sequence 20 from Patent WO02070740. DNA linear PAT 24-MAR-2003
 ACCESSION AX663075
 VERSION AX663075.1 GI:29169369
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 1. .9320
 location/Qualifiers
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Plasmid p5-Puro-CMV- (N'-EGFP)-CMV-Red- (EGFP-EJ)"
 1..1592
 /note="Retroviral Vektor p5NM"
 1617..2216
 /note="unnamed protein product; Puromycin-Resistenzgen aus pRetroon (Clontech, Palo Alto, CA, USA)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80174.1"
 /db_xref="GI:29169370"
 /db_xref="REMBL:CAD80174"
 /translation="MTEYKPTVRLATRDVPRVRLTAAAFADYPATRIHTVDPDRHIE RVTELOELFTVRLDIDGKVVADGAAVAVTTTPEVAGAVFAEIGPRMAELSGSR LAQQQNEGLLAPHRKEPAKELATVGVSDHOGKGLGSANVLPGVEAERAGVPAFL ETAPRNLPYRELGLGTFTADVECPKDRATWCTRPGA"
 2267..2848
 /note="CMV-Promotor aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 2906..>3349
 /note="unnamed protein product; N'-EGFP, abgeleitet von EGFP aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80175.1"
 /db_xref="GI:29169371"
 /db_xref="REMBL:CAD80175"
 /translation="MYSKGEELFTGVVPIVLVDGVDVNGHKFVSVEGEGSDATYKLT LKICITGKLPVWPVPLTTLTVGVOCFSRYDPHMKQHDFFKSNMPEGVQVETIFFK DGGNYKTRAEVKEEGTILVNRILKIDPKEDGNILGHKLEYNVR"
 3374..3392
 /note="Retroviral Vektor p5NM"
 3411..3992
 /note="CMV-Promotor aus p5Red1-N1 (Clontech, Palo Alto, CA, USA)"
 4038..4718
 /note="unnamed protein product; Red aus p5Red1-N1 (Clontech, Palo Alto, CA, USA)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80176.1"
 /db_xref="GI:29169372"
 /db_xref="REMBL:CAD80176"
 /translation="MTRSSKNVTKFMRPKVRMEGTVNGHFEIIEGEGRPYEGHNT VKLVTKGSLPFAWDILSPQYGVSKVYKHPADIPDYKLSFPEGFKEWVNFED GGVVTVDSSLDQDGFIVYKXFGVNFPSDGVFMOKTGMWEASTERLYPRDGLVKG EITHALKLKDGGHYLVEFKSIYNAKKPVQLPGYYVDSKLDITSHEDYITVEQVET

misc_feature 97.6%; Score 661.6; DB 6; Length 9320;
 Best Local Similarity 98.7%; Pred. No. 4.4e-78;
 Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGCTCTCCGAGAACGTCATCACCAGATTCAGGATTCATCGGTTCAAGTGGCGATGGAGGCA 61
 |||||
 Db 4042 TGGCTCTCTCCAGAACGTCATCAGAGGATTCATCGGTTCAAGTGGCGATGGAGGCA 4101
 |||||
 QY 62 CCGTGAAGCGGCCACAGATTCCAGATTCGAGGCGAGGCGAGGCGGCCCTTACGAGGGCC 121
 |||||
 Db 4102 CCGTGAAGCGGCCACAGATTCCAGATTCGAGGCGAGGCGAGGCGGCCCTTACGAGGGCC 4161
 |||||
 QY 122 ACAACACCGTGAAGTTGAAGTGACCAAGGGGGGCCCCCTTCCCTTCCCTTGGGACATCC 181
 |||||
 Db 4162 ACAACACCGTGAAGTTGAAGTGACCAAGGGGGGCCCCCTTCCCTTCCCTTGGGACATCC 4221
 |||||
 QY 182 TGTCCCCCAGTTCCAGTACCGCTCCAAAGGTGTACGTGAAGCACCCTCCGACATCCCGC 241
 |||||
 Db 4222 TGTCCCCCAGTTCCAGTACCGCTCCAAAGGTGTACGTGAAGCACCCTCCGACATCCCGC 4281
 |||||
 QY 242 ACTACAAGAGTGCTCTTCCCCAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAGG 301
 |||||
 Db 4282 ACTACAAGAGTGCTCTTCCCCAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAGG 4341
 |||||
 QY 302 ACGGGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGCAGGCGGCTGCTTCACTACA 361
 |||||
 Db 4342 ACGGGCGGTGGTGACCGTGACCCAGGACTCTCTCCCTGCAGGCGGCTGCTTCACTACA 4401
 |||||
 QY 362 AGGTGAAGTTTCATCGGCGTGAATTCCTCCCGAGCGGCCCTGATGAGAAGAAGACCA 421
 |||||
 Db 4402 AGGTGAAGTTTCATCGGCGTGAATTCCTCCCGAGCGGCCCTGATGAGAAGAAGACCA 4461
 |||||
 QY 422 TGGGTGGAGGCGCTTCCACCGAGCGCTGTACCCCGCGGACGGCTGCTGAGGCGGAGA 481
 |||||
 Db 4462 TGGGTGGAGGCGCTTCCACCGAGCGCTGTACCCCGCGGACGGCTGCTGAGGCGGAGA 4521
 |||||
 QY 482 TCCACAAGCGCCCTGAAGCTGAAGCGAGCGGCGCCACTACCTGGTGGAGTTCAAGTCCATCT 541
 |||||
 Db 4522 TCCACAAGCGCCCTGAAGCTGAAGCGAGCGGCGCCACTACCTGGTGGAGTTCAAGTCCATCT 4581
 |||||
 QY 542 ACATGGCCCAAGAACCGCTGAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGAGCA 601
 |||||
 Db 4582 ACATGGCCCAAGAACCGCTGAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGAGCA 4641
 |||||
 QY 602 TCACCTCCCAACAGAGGACTACACCATCTGGAGCAGTACGAGCGCACCGAGGCGCGCC 661
 |||||
 Db 4642 TCACCTCCCAACAGAGGACTACACCATCTGGAGCAGTACGAGCGCACCGAGGCGCGCC 4701
 |||||
 QY 662 ACCACCTGTTCTGTGA 677
 |||||
 Db 4702 ACCACCTGTTCTGTGA 4717
 |||||

RESULT 10
 AX823860
 LOCUS AX823860
 DEFINITION Sequence 8 from Patent WO03070931. DNA linear PAT 11-DEC-2003
 ACCESSION AX823860
 VERSION AX823860.1 GI:39750176
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 1. .9320
 location/Qualifiers
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Plasmid p5-Puro-CMV- (N'-EGFP)-CMV-Red- (EGFP-EJ)"
 1..1592
 /note="Retroviral Vektor p5NM"
 1617..2216
 /note="unnamed protein product; Puromycin-Resistenzgen aus pRetroon (Clontech, Palo Alto, CA, USA)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80174.1"
 /db_xref="GI:29169370"
 /db_xref="REMBL:CAD80174"
 /translation="MTEYKPTVRLATRDVPRVRLTAAAFADYPATRIHTVDPDRHIE RVTELOELFTVRLDIDGKVVADGAAVAVTTTPEVAGAVFAEIGPRMAELSGSR LAQQQNEGLLAPHRKEPAKELATVGVSDHOGKGLGSANVLPGVEAERAGVPAFL ETAPRNLPYRELGLGTFTADVECPKDRATWCTRPGA"
 2267..2848
 /note="CMV-Promotor aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 2906..>3349
 /note="unnamed protein product; N'-EGFP, abgeleitet von EGFP aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80175.1"
 /db_xref="GI:29169371"
 /db_xref="REMBL:CAD80175"
 /translation="MYSKGEELFTGVVPIVLVDGVDVNGHKFVSVEGEGSDATYKLT LKICITGKLPVWPVPLTTLTVGVOCFSRYDPHMKQHDFFKSNMPEGVQVETIFFK DGGNYKTRAEVKEEGTILVNRILKIDPKEDGNILGHKLEYNVR"
 3374..3392
 /note="Retroviral Vektor p5NM"
 3411..3992
 /note="CMV-Promotor aus p5Red1-N1 (Clontech, Palo Alto, CA, USA)"
 4038..4718
 /note="unnamed protein product; Red aus p5Red1-N1 (Clontech, Palo Alto, CA, USA)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80176.1"
 /db_xref="GI:29169372"
 /db_xref="REMBL:CAD80176"
 /translation="MTRSSKNVTKFMRPKVRMEGTVNGHFEIIEGEGRPYEGHNT VKLVTKGSLPFAWDILSPQYGVSKVYKHPADIPDYKLSFPEGFKEWVNFED GGVVTVDSSLDQDGFIVYKXFGVNFPSDGVFMOKTGMWEASTERLYPRDGLVKG EITHALKLKDGGHYLVEFKSIYNAKKPVQLPGYYVDSKLDITSHEDYITVEQVET

TITLE Methods for conducting site-specific dna recombination
JOURNAL Patent: WO 03070931-A 8 28-AUG-2003;
Vision 7 GmbH (DE)
FEATURES
Location/Qualifiers
source
1. .6893
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Beschreibung der kuenstlichen Sequenz: provirale
Plasmid-DNA; retrovirales Cre-Reporterkonstrukt #Spr#" 601
misc_feature
1. .435
/note="Plasmid-Rueckgrat (pUC)" 661
misc_feature
436. .996
/note="MPSV 5'-LTR (Deltal-31)" 661
primer_bind
997. .1014
/note="PBS (primer binding site)" 661
5'UTR
1015. .1560
misc_feature
1561. .1599
/note="#loxP1#-Sequenz" 661
misc_feature
1600. .2289
/note="DsRed1-Gen aus Discosoma sp." 661
misc_feature
2290. .2343
/note="#loxP2#-Sequenz" 661
misc_feature
2347. .3150
/note="edFP(tag)-Gen aus Aequorea victoria" 661
misc_feature
3154. .4019
/note="#wPRE#" 661
3'UTR
4020. .4080
misc_feature
4081. .4641
/note="3'-LTR" 661
misc_feature
4642. .6893
/note="Plasmid-Rueckgrat (pUC)" 661
ORIGIN
Query Match 97.3%; Score 659.6; DB 6; Length 6893;
Best Local Similarity 98.7%; Pred. No. 8.5e-78;
Matches 665; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGCCTCTCCGAGAACGTCATCACCGAGTTTCATGCGCTTCAAGGTTCGCGATCGAGGGCA 61
DB 1604 TGGCCTCTCCGAGAACGTCATCACCGAGTTTCATGCGCTTCAAGGTTCGCGATCGAGGGCA 1663
QY 62 CCGTGAACCGCCAGGTTGAGATCGAGGCGGAGGCGGCGCCCTACGAGGGCC 121
DB 1664 CCGTGAACCGCCAGGTTGAGATCGAGGCGGAGGCGGCGCCCTACGAGGGCC 1723
QY 122 ACAACACCGTGAAGTTGAAGTGAACGAGGCGGCGCCCTCGCCCTTGGCGGACATCC 181
DB 1724 ACAACACCGTGAAGTTGAAGTGAACGAGGCGGCGCCCTCGCCCTTGGCGGACATCC 1783
QY 182 TGTCCCCCGAGTTCCAGTA CGGCTCCAGAGTTGATGAGTGAAGCACCCTCGGACATCCCCG 241
DB 1784 TGTCCCCCGAGTTCCAGTA CGGCTCCAGAGTTGATGAGTGAAGCACCCTCGGACATCCCCG 1843
QY 242 ACTACAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGAGCGGCGGTGATGAATCTTCGAGG 301
DB 1844 ACTACAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGAGCGGCGGTGATGAATCTTCGAGG 1903
QY 302 ACGGCGGCGTGGCGACCGTGAACGAGTCTCTCCCTGCGAGGAGCGGCTGTTCATCTACA 361
DB 1904 ACGGCGGCGTGGTGAACGAGTCTCTCCCTGCGAGGAGCGGCTGTTCATCTACA 1963
QY 362 AGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGAGGCGGCGGTGATGACAGAGAGACCA 421
DB 1964 AGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGAGGCGGCGGTGATGACAGAGAGACCA 2023
QY 422 TGGGCTGGAGGCGCTTCCACCGAGGCGCTGTACCCCGCGGAGCGGCGTGTGAAGGGCGAGA 481
DB 2024 TGGGCTGGAGGCGCTTCCACCGAGGCGCTGTACCCCGCGGAGCGGCGTGTGAAGGGCGAGA 2083
QY 482 TCCACAGGCGCTGAAGCTGAAGAGCGGCGGCGCTACCTCGTGGAGTTCAAGTCCATCT 541
DB 2084 TCCACAGGCGCTGAAGCTGAAGAGCGGCGGCGCTACCTCGTGGAGTTCAAGTCCATCT 2143

QY 542 ACATGGCCAAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACACCAAGCTGGACA 601
DB 2144 ACATGGCCAAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACTCCAAAGCTGGACA 2203
QY 602 TCACCTCCCAACAGGAGCTACACATCGTGGAGAGTACGAGCGACCGAGCGCGCCGCC 661
DB 2204 TCACCTCCCAACAGGAGCTACACATCGTGGAGAGTACGAGCGACCGAGCGCGCCGCC 2263
QY 662 ACCACCTGTTCTCTG 675
DB 2264 ACCACCTGTTCTCTG 2277
RESULT 11
AY342347 10141 bp DNA linear SYN 21-OCT-2003
LOCUS Red H-Pelican DsRed.T4 transformation vector, complete sequence.
DEFINITION Red H-Pelican DsRed.T4 transformation vector
ACCESSION AY342347
VERSION AY342347.1 GI:33358309
KEYWORDS
SOURCE
ORGANISM
Red H-Pelican DsRed.T4 transformation vector
Red H-Pelican DsRed.T4 transformation vector
artificial sequences; vectors.
REFERENCE
1 (bases 1 to 10141)
Barolo, S. and Posakony, J.W.
Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP
[DsRed.T4] and Nuclear RFP in Insulated Vectors
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 10141)
Barolo, S. and Posakony, J.W.
Direct Submission
AUTHORS
TITLE
Submitted (14-JUL-2003) Division of Biology/CDB, UCSD, 4121 Bonner
Hall, MC 0349, La Jolla, CA 92093-0349, USA
JOURNAL
LOCATION/Qualifiers
1. .10141
/organism="Red H-Pelican DsRed.T4 transformation vector"
/mol_type="other DNA"
/db_xref="taxon:240544"
FEATURES
source
ORIGIN
Query Match 95.8%; Score 649.8; DB 12; Length 10141;
Best Local Similarity 97.5%; Pred. No. 1.5e-76;
Matches 660; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATGCGCTTCAAGGTTCGCGATCGAGGGC 60
DB 285 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATGCGCTTCAAGGTTCGCGATCGAGGGC 344
QY 61 ACCGTGAACCGCCAGGTTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 345 TCGGTGAACCGCCAGGTTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 404
QY 121 CACACACCGTGAAGTTGAAGTGAACGAGTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGG 180
DB 405 ACCGAGCGCCAAAGCTGAAGTGAACGAGTTCGAGATCGAGGCGGCGGCGGCGGCGGCGGCGG 464
QY 181 CTGTCCCCCGAGTTCCAGTACCGCTTCCAGGTGATGAGTGAAGCACCCTCGGACATCCCC 240
DB 465 CTGTCCCCCGAGTTCCAGTACCGCTTCCAGGTGATGAGTGAAGCACCCTCGGACATCCCC 524
QY 241 GACTACAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGAGCGGCGGTGATGAATCTCGAG 300
DB 525 GACTACAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGAGCGGCGGTGATGAATCTCGAG 584
QY 301 GACGCGGCGTGGCGACCGTGAACCGAGGACTCTCTCCCTGCGAGGCGGCTGTTCATCTAC 360
DB 585 GACGCGGCGTGGTGAACCGTGAACCGAGGACTCTCTCCCTGCGAGGCGGCTGTTCATCTAC 644
QY 361 AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 420
DB 645 AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 704

```

QY 421 ATGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTCTCAAGGGCGAG 480
Db 705 ATGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTCTCAAGGGCGAG 764
QY 481 ATCCACAGGCGCTCAAGCTGAAGACGGCGGCGCACTACTCTGGTGGAGTTCAAGTCCATC 540
Db 765 ATCCACAGGCGCTCAAGCTGAAGACGGCGGCGCACTACTCTGGTGGAGTTCAAGTCCATC 824
QY 541 TACATGGCCCAAGACCGCTGAGCTGCGCGGCTACTACTACGTGGACCAACCAAGCTGGAC 600
Db 825 TACATGGCCCAAGACCGCTGAGCTGCGCGGCTACTACTACGTGGACCAACCAAGCTGGAC 884
QY 601 ATCACTCCCAAGACGAGCTACACCATCTGTGAGAGTAGTACGAGGCGGCGGCGGCGC 660
Db 885 ATCACTCCCAAGACGAGCTACACCATCTGTGAGAGTAGTACGAGGCGGCGGCGGCGC 944
QY 661 CACCACCTGTTCTCTGTA 677
Db 945 CACCACCTGTTCTCTGTA 961

RESULT 12
AY342348 10276 bp DNA linear SYN 21-OCT-2003
LOCUS Red H-Stinger DsRed.T4-NLS transformation vector, complete
DEFINITION
ACCESSION AY342348
VERSION AY342348.1 GI:33359310
KEYWORDS
SOURCE Red H-Stinger DsRed.T4-NLS transformation vector
ORGANISM Red H-Stinger DsRed.T4-NLS transformation vector
REFERENCE 1 (bases 1 to 10276)
AUTHORS Barolo, S. and Posakony, J.W.
TITLE Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP
[DsRed.T4] and Nuclear RFP in Insulated Vectors
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10276)
AUTHORS Barolo, S. and Posakony, J.W.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2003) Division of Biology/CDB, UCSD, 4121 Bonner
Hall, MC 0349, La Jolla, CA 92093-0349, USA
FEATURES
source
1. 10276
/organism="Red H-Stinger DsRed.T4-NLS transformation
vector"
/mol_type="other DNA"
/db_xref="taxon:240545"

ORIGIN
Query Match 95.8%; Score 649.8; DB 12; Length 10276;
Best Local Similarity 97.5%; Pred. No. 1.5e-76;
Matches 660; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 ATGGCTCTCCGAGACGTCATCACCGAGTTTCATGGCTTCAAGTGGCGCATGGAGGC 60
Db 288 ATGGCTCTCCGAGACGTCATCACCGAGTTTCATGGCTTCAAGTGGCGCATGGAGGC 347
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGG 120
Db 348 TCCGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGG 407
QY 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 408 ACCCAGACCGCAAGCTGAAGGTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 467
QY 181 CTGTCCCGCCAGTTCACGTCAGCTCCAGGTGTCAGTGAAGCACCCCGCGCATCCCC 240
Db 468 CTGTCCCGCCAGTTCACGTCAGCTCCAGGTGTCAGTGAAGCACCCCGCGCATCCCC 527
QY 241 GACTACAAGAGCTGTCCTTCCCGGAGGCTTCAAGTGGAGCGCGTGAATCACTTCGAG 300
Db 528 GACTACAAGAGCTGTCCTTCCCGGAGGCTTCAAGTGGAGCGCGTGAATCACTTCGAG 597

```

```

QY 301 GACGGCGGCTGGCGACCGTGAACCGAGTCTCTCCCTGCAGACGGCTGCTTCATCTAC 360
Db 588 GACGGCGGCTGGCGACCGTGAACCGAGTCTCTCCCTGCAGACGGCTGCTTCATCTAC 647
QY 361 AAGGTGAAGTTTCATCGGCTGAACCTTCCTCCGACGGCGGCTGATGACGAAGAAGACC 420
Db 648 AAGGTGAAGTTTCATCGGCTGAACCTTCCTCCGACGGCGGCTGATGACGAAGAAGACT 707
QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGAGCGGCTGTGAAGGGCGAG 480
Db 708 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGAGCGGCTGTGAAGGGCGAG 767
QY 481 ATCCCAAGGCGCTGAAGCTGAAGGACGGCGGCGGCTACCTGGTGGAGTTCAAGTCCATC 540
Db 768 ATCCCAAGGCGCTGAAGCTGAAGGACGGCGGCGGCTACCTGGTGGAGTTCAAGTCCATC 827
QY 541 TACATGGCCCAAGACCGCTGAGCTGCGCGGCTACTACTAGTGGACCAACCAAGCTGGAC 600
Db 828 TACATGGCCCAAGACCGCTGAGCTGCGCGGCTACTACTAGTGGACCAACCAAGCTGGAC 887
QY 601 ATCACTCCCAAGACGAGCTACACCATCTGTGAGAGTAGTACGAGGCGGCGGCGGCGC 660
Db 888 ATCACTCCCAAGACGAGCTACACCATCTGTGAGAGTAGTACGAGGCGGCGGCGGCGC 947
QY 661 CACCACCTGTTCTCTGTA 677
Db 948 CACCACCTGTTCTCTGTA 964

RESULT 13
AF506025 681 bp DNA linear SYN 19-JUN-2002
LOCUS Synthetic construct dimeric red fluorescent protein gene, complete
DEFINITION
ACCESSION AF506025
VERSION AF506025.1 GI:21464833
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 681)
AUTHORS Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S.,
Zacharias, D.A., and Tsien, R.Y.
TITLE A monomeric red fluorescent protein
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7877-7882 (2002)
MEDLINE 22056088
PUBMED 12060735
REFERENCE 2 (bases 1 to 681)
AUTHORS Campbell, R.E. and Tsien, R.Y.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La
Jolla, CA 92093, USA
FEATURES
source
1. 681
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="from Discosoma sp.; Clontech vector pDsRed-N1 with
human optimized codon usage"
1. 681
/note="dimer2; engineered variant of tetrameric red
fluorescent protein from Discosoma deposited in GenBank
Accession Number AAF03369"
/codon_start=1
/transl_table=1
/product="dimeric red fluorescent protein"
/protein_id="AA054542.1"
/db_xref="GI:21464834"
/transl_table="MVASSEDVKEFMFRKVRMEGVSNGHEFIEGEGEGRPYEGTQT
AKLVKGGPLPFAWDILSPQFQVSKAYVHPADIPDYKLSFPFGFKVWVNFED
GGVTVTQDSLSQDGLIYKVRFGTNPDPGVPWOKTGWSEASTERLYPRDGLKG
EIHQALKLKGHLYVEFTIYMAKPVQLPGYVYVDTKLDITSHEDYITVQYERS

```


PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source

12060735
 2 (bases 1 to 1395)
 Campbell,R.E. and Tsien,R.Y.
 Direct Submission
 Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La
 Jolla, CA 92093, USA
 Location/Qualifiers
 1. .1395
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="from *Drosophila* sp.; Clontech vector pDsRed-N1 with
 human optimized codon usage"
 1. .1395
 /note="tdimer2(12); engineered variant of tetrameric red
 fluorescent protein from *Drosophila* deposited in GenBank
 Accession Number AF03369"
 /codon_start=1
 /transl_table=11
 /product="tandem-dimer red fluorescent protein"
 /protein_id="AA054543.1"
 /db_xref="GI:21464836"
 /translation="MVASSEIVIEKFMRFKVRMEGVSNGHFEIIEGEGRGREYEGTQT
 AKLVTKGGPLPFPAFDILSPQYGSKAYVKHPADIPDYKKLSFEGFKWERVMFED
 GGVTVTDDSSLDGTLIIYKVRFGTNPFPDGPVQKMTMGWEASTERLYPRDGLVKG
 EIHQALKDGGHLYVQIYMAKKPVQLPQYGVYVTDKLDITSNEDYIIIEQVZRS
 EGRHLFTGHGTGTGSGSSATGSEDIYKEFMRFKVRMEGVSNGHFEIIEGEGRGPR
 YEGTQAKLVTKGGPLPFPAFDILSPQYGSKAYVKHPADIPDYKKLSFEGFKWER
 VMNFEDGGVTVTDDSSLDGTLIIYKVRFGTNPFPDGPVQKMTMGWEASTERLYPR
 DGLVKGELHQALKDGGHLYVFEKTIYMAKKPVQLPQYGVYVTDKLDITSNEDYIY
 EQERSEGRHHFLF"
 Query Match 92.6%; Score 627.6; DB 12; Length 1395;
 Best Local Similarity 95.7%; Pred. No. 1.9e-73;
 Matches 645; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
 2 TGGCTCTCCGAGAACGTCATCCAGTGTTCAATCGCTTCAAGTGTCGCATGGAGGCA 61
 5 TGGCTCTCTCCGAGGACGTCATCAAGAGGTTTCATCGCTTCAAGTGTCGCATGGAGGCT 64
 62 CCGTGAAAGGGCCACAGTTCGAGTCGAGGGCGAGGGCGGCCGCCCTACGAGGGCC 121
 65 CCGTGAAAGGGCCACAGTTCGAGTCGAGGGCGAGGGCGGCCGCCCTACGAGGGCA 124
 122 ACAACACCGTGAAGTTGAAGTTGACCAAGGGGGGGCCCCCTGCGCCCTTCGCGTGGGACATCC 181
 125 CCGAGCCGCCAAGCTGAAGTTGACCAAGGGCGGGCCCCCTGCGCTTCGCTGGGACATCC 184
 182 TGTCTCCCCCAGTTCACGATACGGCTCCAAAGGTGTACGTGAAGCAACCCCGCCGACATCCCCG 241
 185 TGTCTCCCCCAGTTCACGATACGGCTCCAAAGGTGTACGTGAAGCAACCCCGCCGACATCCCCG 244
 242 ACTACAAGAGGTGTCCTTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACCTTCGAGG 301

ORIGIN

302	ACGGCGGGTGGCGACCGTGACCCAGGACTCTCCCTGCAGACGGCTGTTTCATCTACA	361
305	ACGGCGGGCTGGTGACCGTGACCCAGGACTCTCCCTGCAGACGGCACGCTGATCTACA	364
362	AGGTGAAGTTTCATCGGCGTGAACTTCCCTCCGACGGCCCCGCTGATGCAGAAAGAACCA	421
365	AGTGAAGTTCCGCGGGCACCACTTCCCCCGACGGCCCCGTAAATGCAGAGAAACCA	424
422	TGGGTGGGAGGCTTCAACCGAGCGCTGTATCCCGCGGACGGCGTCTGAAGGGCGAGA	481
425	TGGGTGGGAGGCTTCAACCGAGCGCTGTATCCCGCGGACGGCGTCTGAAGGGCGAGA	484
482	TCCCAAGGCCCTGAAGCTGAAGGACGGCGGCACCTACCTGGTGGAGTTTCAAGTCCATCT	541
485	TCCACAGGCCCTGAAGCTGAAGGACGGCGGCACCTACCTGGTGGAGTTTCAAGACCATCT	544


```
QY 542 ACATGGCCCAAGACCGCTGAGCTGCGCGCTACTACTAGTGGACACCAAGCTGGACA 601
Db 545 ACATGGCCCAAGACCGCTGAGCTGCGCGCTACTACTAGTGGACACCAAGCTGGACA 604
QY 602 TCACCTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGCGCGCC 661
Db 605 TCACCTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCTCCGAGCGCGCC 664
QY 662 ACCACCTGTTCTCTG 675
Db 665 ACCACCTGTTCTCTG 678

RESULT 15
AX233584
LOCUS AX233584 681 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 8 from Patent WO0162919.
ACCESSION AX233584
VERSION AX233584.1 GI:15593307
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
          artificial sequences.
REFERENCE
1 Nelson,D., Zamaira,E. and Tsien,R.
  AUTHORS Modified fluorescent proteins
  TITLE Patent: WO 0162919-A 8 30-AUG-2001;
  JOURNAL Aurora Biosciences Corporation (US)
FEATURES
source
1..681
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Mutant Anthozoan red fluorescent protein"
1..681
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC69734.1"
/db_xref="GI:15593308"
/db_xref="RENTREMBL:CAC69734"
/translation="MVRSSKNVKEFMFKVMEGTUNGHEFIEGEGSRPYEGHNT
VKLVTKGGPLFPAMDILSPQYIGSKVYVHPADIPDYKKLSFPEGFKWRVWNPED
GGVVTVDQSLQDGFYKVFYGVNPSDPVQKMGWEASTERYLPDRGVKLG
EIKHKLKDKDGHYLVBFKSIYMAKRPVQLPGYYVVDKSLDITSHNEDYIVEQVET
EGRHHLFL"

CDS
Query Match 90.6%; Score 614.2; DB 6; Length 681;
Best Local Similarity 95.1%; Pred. No. 1.3e-71;
Matches 634; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 12 CGAGAACGTGATCATCCAGGTTTCATGCGCTTCAAGTGGCGATGGAGGGACCGCTGAACGG 71
Db 15 CAAGACGTGATCAAGAGTTCATAGGTTCAAGTGGCGATGGAGGGACCGCTGAACGG 74
QY 72 CCACGAGTTCAGATTCAGAGGCGAGGGCGGCGCCCTTACGAGGGCCACCAACCGT 131
Db 75 CCACGAGTTCAGATTCAGAGGCGAGGGCGGCGCCCTTACGAGGGCCACCAACCGT 134
QY 132 GAAGTTGAAGTGACCAAGGCGGCGCCCTTCCGCTTGGGACATCTGTCCCCCA 191
Db 135 GAAGCTTAAGTGACCAAGGCGGCGCCCTTCCGCTTGGGACATCTGTAGCCCCCA 194
QY 192 GTTCCAGTACGGCTCCAGAGTGTACGTGAAGACACCCCGCGACATCCCGACTACAAGAA 251
Db 195 GTTCCAGTACGGCAGCAAGGTGTACGTGAAGACACCCCGCGACATCCCGACTACAAGAA 254
QY 252 GCTGCTCTCCCGAGGGCTTCAGTGGAGCGCGTGAATCACTTCGAGACCGCGCGT 311
Db 255 GCTGAGCTTCCCCGAGGGCTTCAGTGGAGAGGGTGAATCACTTCGAGACCGCGCGT 314
QY 312 GGCACCGGTGACCCAGGACTCTCTCCCTGCAGGACGGCTCTCATCTACAAGGTGAAGTT 371
```

```
Db 315 GGTGACCGTGACCCAGGACAGAGCTGCAGACGGCTTTCATCTACAGGTGAAGTT 374
QY 372 CATCGCGGTGAACATTCCCTCCGACGGCCCGCTGATGAGAGAGAACCAATGGGCTGGGA 431
Db 375 CATCGCGGTGAACATTCCCTCCGACGGCCCGCTGATGAGAGAGAACCAATGGGCTGGGA 434
QY 432 GGCCTCCACCGAGCGCTGTACCCCGGACGGCGTCTGAAGGCGGAGATCCACAAAGC 491
Db 435 GGCCTCCACCGAGCGCTGTACCCCGGACGGCGTCTGAAGGCGGAGATCCACAAAGC 494
QY 492 CCTGAAGCTGAAGGACGCGCGCCACTACTCTGTTGGAGTTCAAGTCCATCTACATGGCAA 551
Db 495 CCTGAAGCTGAAGGACGCGCGCCACTACTCTGTTGGAGTTCAAGTCCATCTACATGGCAA 554
QY 552 GAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACCTCCA 611
Db 555 GAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACATCACCAAGCTGGACATCACCAAGC 614
QY 612 CAAAGGAGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGCGCGCCACCACTGTT 671
Db 615 CAAAGGAGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGCGCGCCACCACTGTT 674
QY 672 CCTGTAA 678
Db 675 CCTGTGA 681
```

Search completed: August 3, 2004, 14:04:16
Job time : 4537.5 secs

Sequence 12, Appl
Sequence 17, Appl
Sequence 13, Appl
Sequence 18, Appl
Sequence 12, Appl
Sequence 17, Appl
Sequence 9, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 30, Appl
Sequence 33, Appl
Sequence 33, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 31, Appl
Sequence 21, Appl

Sequence 12, Appl
Sequence 17, Appl
Sequence 13, Appl
Sequence 18, Appl
Sequence 12, Appl
Sequence 17, Appl
Sequence 9, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 30, Appl
Sequence 33, Appl
Sequence 33, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 31, Appl
Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342372
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, Iii, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

Query Match 64.6%; Score 438; DB 4; Length 678;
Best Local Similarity 77.9%; Pred. No. 3.8e-76;
Matches 528; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 1 ATGGCTCTCCGAGACGTCATCCAGGTTTCAGGTTCAAGGTCGCGATCGAGGC 60
DB 1 ATGAGGTTTCCAAAGATGTTATCAAGAGTTCAAGGTTTAAGTTTCGATGAAGGA 60
QY 61 ACCGTGAACGCCACGAGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCTTACGAGGC 120
DB 61 ACCGTGAACGCCACGAGTTTGAATAGAGGCGGAGGCGGAGGCGCCATCGAGGC 120
QY 121 CACACACCGTGAAGTTGAAGTGCACGAGGCGGCGCCCTTCCCTTGGGACATC 180
DB 121 CACATACCGTGAAGTTGAAGTGCACGAGGCGGCGCCCTTCCCTTGGGATAT 180
QY 181 CTGTCCCGCCAGTTCCAGTACGCTCCAGGTTTACGTAAGCAGCCCGCGCATCCCC 240
DB 181 TTGTCCACCAATTTAGTATGGAAGCAGGATATATGTCAAGCAGCCCTTCCGACATCA 240
QY 241 GACTACAAAGAGCTGTCTTCCCGAGGCTTCAAGTGGGAGCGGCTGATGAATTCGAG 300
DB 241 GACTATAAAGAGCTGTCTTCTTCAAGGATTTAAATGGGAAGGCTCATGAATCTTGA 300

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 11:04:09 ; Search time 68.5343 Seconds
(without alignments)
5490.050 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggctctccgagacgt.....gccaccacgttctctgttaa 678

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
6: /cgn2_6/prodata/2/ina/5B_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	438	64.6	678	4	US-09-459-956-6
2	257.4	38.0	699	4	US-09-459-956-5
3	251.8	37.1	720	4	US-09-839-650-1
4	216.6	31.9	801	4	US-09-459-956-7
5	144.8	21.4	690	4	US-09-459-956-2
6	119	17.6	1104	3	US-09-277-716-30
7	119	17.6	1104	4	US-09-609-161B-30
8	119	17.6	1279	3	US-09-277-716-31
9	119	17.6	1279	4	US-09-609-161B-31
10	110.4	16.3	696	4	US-09-459-956-4
11	100.2	14.8	717	4	US-09-214-909-1
12	99.2	14.6	696	4	US-09-459-956-3
13	99	14.6	717	4	US-09-486-241-31
14	98.4	14.5	1079	4	US-09-609-161B-15
15	98.4	14.5	1079	4	US-09-626-581B-64
16	98.4	14.5	1079	4	US-09-415-765B-64
17	98.4	14.5	1079	4	US-09-626-580C-64
18	98.4	14.5	1085	3	US-09-277-716-15
19	95.2	14.0	1021	4	US-09-839-650-2
20	94.4	13.9	782	1	US-08-532-390-40
21	94.4	13.9	782	3	US-08-717-294-40
22	93	13.7	720	3	US-09-172-063-14
23	93	13.7	720	4	US-09-316-919-15
24	93	13.7	720	4	US-09-602-641-14
25	92.8	13.7	7353	4	US-08-786-531B-1
26	91.4	13.5	717	4	US-09-513-783A-43
27	91.4	13.5	720	3	US-09-094-359-9

QY 301 GACGGCGCGTGGCGACCGTGAACCCAGGACTCCCTCCCTGAGGACGGCTGCTTCACTAC 360
DB 301 GACGGTGGCGTGTCTGTAACCCAGANTCCAGTTTGAGATGGCTGTTTCACTAC 360
QY 361 AAGGTGAAGTTCATCGCGGTGAATCTCCCTCCGACGGCCCCGATGATGCAAGAGAC 420
DB 361 AAGGTGAAGTTCATCGCGGTGAATCTCCCTCCGACGGCCCCGATGATGCAAGAGAC 420
QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAG 480
DB 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAG 480
QY 481 ATCCACAGGCGCTGAAGCTGAAGGACGGCGGCGCACTACCTGTGTGAGTTCAAGTCCATC 540
DB 481 ATTCAATAGGCTCTGAAGCTGAAGGACGGCGTGTACCTAGTTGAATTCAAAAGTATT 540
QY 541 TACATGGCAAGAGCCGCTGACGCTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
DB 541 TACATGGCAAGAGCCGCTGACGCTCCCGGCTACTACTAGTGGACACCAAGCTGGAT 600
QY 601 ATACCTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660
DB 601 ATACAGGACCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGACG 660
QY 661 CACCACCTGTTCCTGTAA 678
DB 661 CACCACCTGTTCCTTTAA 678

RESULT 2
US-09-459-956-5
; Sequence 5, Application US/0945956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

Query Match 38.0%; Score 257.4; DB 4; Length 699;
Best Local Similarity 62.4%; Pred. No. 2.1e-41;
Matches 421; Conservative 0; Mismatches 251; Indels 3; Gaps 1;

QY 1 ATGGCCCTCTCCGAGAACGTCAACCGAGTTTCATGGGTTCAAGGTGCGCATGGAGGC 60
DB 1 ATGAGTTGTTCCAAAGAGTGTGATCAAGAGAAATGTTGATCGATCTTCATCTGGAAGA 60
QY 61 ACCGTGAACCGCCACAGTTCGAGATCGAGGCGGAGGCGCGCCCTACGAGGC 120
DB 61 ACGTTCATGGGCACTACTTTGAAATAAAGGCAAGAGAAAGACAGCTTATGANGC 120
QY 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGGGGGCGCCCTCGCCCTTCGCTGGGACATC 180
DB 121 ACCAATACCGTCAAGCTCGAGGTTACCAAGGGTGGACCTCTGCCATTTGTTGGCATATT 180
QY 181 CTGTCCCCCAGTTCAGTACGCTCCAGGTGTAGTGAAGCACCCCGCGGACATCCCC 240

DB 181 TTGTGCCCAATTTTCAGTATGGAACAAGGCATTTGTCCACCACCCTGACACATCAT 240
QY 241 GACTACAGAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATCTGAG 300
DB 241 GATTAATCTAAGAGCTGTCTCTCCCGAGGGATATACATGGGAACGGTCCATGCACTTTGAA 300
QY 301 GACGGCGCGTGGCGACCGGTGACCCAGGACTCCCTCCCTGACAGGACGGCTGCTTCACTAC 360
DB 301 GACGGTGGCTTGTGTGTATCACCAGTATATCAGTTTGACAGGCAACTGTTTCTACTAC 360
QY 361 AAGGTGAAGTTCATCGCGGTGAATCTCCCTCCGACGGCCCCGATGATGCAAGAGAC 420
DB 361 GACATCAAGTTCATCGCGGTGAATCTCCCTCCGACGGCCCCGATGATGCAAGAGAC 420
QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAG 480
DB 421 ACTGGCTGGAGACCGAGGACCTGAGGGTTTGTATCTCTGTGATGTTGTTGNTAGAGAC 480
QY 481 ATCCACAGGCGCTGAAGCTGAAGGACGGCGGCGCACTACCTGTGTGAGTTCAAGTCCATC 540
DB 481 ATCCATCATGCTCTGACAGTTGAAGGAGTGTGCTATTACGCACTGACANTTAAACTGTT 540
QY 541 TACATGGCAAGAGCCGCTGACGCTCCCGGCTACTACTAGTGGACACCAAGCTG 597
DB 541 TACAGGGCCAAAGAGGGCGCTTGAAGATGCCAGGGTATCACTATGTTGACCCAAACTG 600
QY 598 GACATCACCTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGC 657
DB 601 GTTATATGGAACAACGACAAGAAATTCATGAAGTTGAGGAGCATGAATCCCGGTTGCA 660
QY 658 GCCCACCACCTGTTTC 672
DB 661 CGCCACCACCTCGTTC 675

RESULT 3
US-09-839-650-1
; Sequence 1, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Humanized R. mulleri polynucleotide
; NAME/KEY: misc feature
; LOCATION: (1)-(720)
; OTHER INFORMATION: Humanized DNA sequence
US-09-839-650-1

Query Match 37.1%; Score 251.8; DB 4; Length 720;
Best Local Similarity 62.4%; Pred. No. 2.6e-40;
Matches 394; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 28 GAGTTCATGCGCTTCAAGTTCGCGATCGAGGCGACCGGTGAACGCCACCGAGTTCGAGATC 87
DB 40 GAGGTGATGAGCTACAAGGTGAACCTGGAGGGATCTGTGAACCAACCAACGTTTCCCATG 99
QY 88 GAGGGCGAGGCGAGGGCGCGCCCTTACGAGGGCCACCAACCCGTGAAGTTGAAGGTGACC 147
DB 100 GAGGGCTCGGCAAGGCGCAACATCTGTTCCGCAACACAGCTGTTGTCAGATCCCGGTGACC 159
QY 148 AAGGGCGGCGCCCTTGCCTTCGCTGGGACATCTGTCTCCCGCAGTTCAGTACGGCTCC 207

Db 61 TGTGTAATGGGATTACTTTACCGTCAAGGTGAAGGCAACGGGAAGCCATACGAAGGG 120
 QY 121 CACAAACCGTGAAGTTGAAGTGCACCAAGGC-----GGCCCCCTGGCCCTGGCTGG 174
 Db 121 ACGCAGACTTCGACTTTTAAAGTACCATGCGCAAGCGTGGGCCCTTGCATTCTCTTT 180
 QY 175 GACATCTGTCCCCCAGTTCCTAGTACGGTCCAAAGGTGATCGTGAAGCACCCCGCGAC 234
 Db 181 GACATATCTACAGTGTTCAAATATGAAATCGATGCTTTACTGCTATCTCCACGT 240
 QY 235 ATCCCGCACTACAAAGTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGTGAAC 294
 Db 241 ATGCCCGACTATTTCAAACCAAGCATTTCTGACGGAATGTCATGAAAGGACTTTTACC 300
 QY 295 TTCAGGACCGCGCGTGGCGACCGTGAACCCAGGACTCTCTCCCTGAGGACGGCTCTTC 354
 Db 301 TATGAAGATGGAGAGTTGCTACAGCCAGTTGGGAATAAGCCTTAAAGGCACTGCTTT 360
 QY 355 ATCTAAGGTGAAGTTTCATCGCGTGAATTCCTCCCTCGACGGCCCCGTGATGCAAG 414
 Db 361 GAGCACAATCCAGTTTTCATGAGTGAATTTCTCTGCTGATGACCTGTGATGGCGAAG 420
 QY 415 AAGACCATGCGTGGGAGGCTCCACCGAGCGCTGTACCCCGCGACGCGGTGCTGAAG 474
 Db 421 AAGACAATGTTGGGACCATCTTTTGAGAAATGACTGTCTGCGATGGAATATTGAAG 480
 QY 475 GGGAGATCCACAAGGCCCTTGAAGTGAAGAGCGGGCGCACTACTGTTGAGTTCAG 534
 Db 481 GGTGATGTCACCGGTTTCTCATGCTGCAAGGAGTGGCAATTCAGATGCCAATCCAC 540
 QY 535 TCCATCTACATGCCAGAGACCGGTGACGTGCGCGGTACTACTACGTGGACACCAAG 594
 Db 541 ACTTCTTACAGACAAAGAACCGGTGACATGCCACCAACCATGTGTGGAAACATGCG 600
 QY 595 CTGACATCACTCCCAACAGGAGGACTACACCATGTTGAGCAGTACGAGGCGCACCG 652
 Db 601 ATTGGAGGACCGACCTTGACAAAGGTGGCAACAGTTCAGCTGACGAGGACCGTGG 658

RESULT 6

US-09-277-716-30
 ; Sequence 30, Application US/09277716A
 ; Patent No. 6232107

GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce
 ; APPLICANT: Szent-Gyorgyi, Christopher
 ; APPLICANT: PROLUME, LTD.
 ; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
 ; CURRENT APPLICATION NUMBER: US/09/277,716A
 ; CURRENT FILING DATE: 1999-03-26
 ; EARLIER APPLICATION NUMBER: 60/102,939
 ; EARLIER FILING DATE: 1998-10-01
 ; EARLIER APPLICATION NUMBER: 60/089,367
 ; EARLIER FILING DATE: 1998-06-15
 ; EARLIER APPLICATION NUMBER: 60/079,624
 ; EARLIER FILING DATE: 1998-03-27
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 30

LENGTH: 1104

TYPE: DNA

ORGANISM: Ptilosarcus gurneyi

FEATURE:

NAME/KEY: CDS

LOCATION: (34)..(747)

FEATURE:

OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)

US-09-277-716-30

Query Match

Best Local Similarity 17.6%; Score 119; DB 3; Length 1104;

Matches 308; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

QY 23 TCACCGAGTTTCATCGGCTTCAAGGTGGCGCATGGAGGCGACCGTGAACGCGCAGGTTGCG 82
 Db 65 TGAAGAGATTATGTCGCGAAAGCTAGCGTTGAAGGAATCGTGAAACAATCACGTTTTTT 124
 QY 83 AGATCGAGGCGAGGCGAGGCGCGCCCTACGAGGCGCAACACACCGTGAAGTTGAAGG 142
 Db 125 CCATGGAAGGATTTTGGAAAAGGCAATGTATTTATTTGGAAAACCAATTGATGCAATCCGG 184
 QY 143 TGACCAAGGCGCGCCCTTGCCTTCGCTTGGGACATCCTGTCCCGCCAGTTCAGTACG 202
 Db 185 TTACAAAGGAGGTCCTTGGCATTCGCTTTCGATATTTGTTCCATAGCTTTTCAATACG 244
 QY 203 GCTCCAAAGTGTAGTGAAGCACCCCGCGGACATCCCGGACTCAAGAAAGTGTCTCTTC 262
 Db 245 GGAATCGCACTTTCACGAAATATCCAGACGACATTCGGGACTACTTTGTTCAATCATTC 304
 QY 263 CCGAGGCGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGGACGGCGGTGGCGACCGTGA 322
 Db 305 CGGTGGAATTTTCTACGAAAGAAATCTACGCTTTGAAGATGGCGCCATTGTTGACATTC 364
 QY 323 CCCAGGACTCTCCCTCGAGGACGGCTGCTTCATCTTCAAGGTGAAGTTTCAATGGCGTGA 382
 Db 365 GTTCAGATATTAAGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAACG 424
 QY 383 ACTTCCCTCGAGCGCGCGGTGATGACAGAGACCATGGCTGGGAGGCGCTCCACCG 442
 Db 425 GTTTCCTTAGTAAACGACCGGTGATGCAAAAGCCATCTCTCGGATGGAGCCATCGTTG 484
 QY 443 AGCGCTGTATACCCCGCGACCGGCTGTGAAGGGCGAGATCCCAAGGCCCTTGAAGTGA 502
 Db 485 AGTGGTCTACATGAACAGCGGCGTCTGTTGGGCGAAGTAGATCTGTTTACAACTCG 544
 QY 503 AGGAGCGGGCGCACTACTGTTGGAGTTCAAGTCCATCTACATGGCCCAAGAGCCCGTGC 562
 Db 545 AGTCAGGGAACATTAATCTCGTGCACATGAAAGCTTTTACAGATCCAAAGGTGGAGTGA 604
 QY 563 AGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACCTCCCAACACGAGACT 622
 Db 605 AAGAAATCCCGAATATCATCTTATCCATCATCTGCTGGAGAAACCTTACGTGGAAGAG 664
 QY 623 ACACCATGTTGACGACGATACGAG 645
 Db 665 GAAGCTTCGTGGAACAACAGAG 687

RESULT 7

US-09-609-161B-30

; Sequence 30, Application US/09609161B

; Patent No. 6436682

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC

; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG

; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

; FILE REFERENCE: 24729-121B

; CURRENT APPLICATION NUMBER: US/09/609,161B

; CURRENT FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: 09/277,716

; PRIOR FILING DATE: 1999-03-26

; PRIOR APPLICATION NUMBER: 60/102,939

; PRIOR FILING DATE: 1998-10-01

; PRIOR APPLICATION NUMBER: 60/089,367

; PRIOR FILING DATE: 1998-06-15

; PRIOR APPLICATION NUMBER: 60/079,624

; PRIOR FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 30

; LENGTH: 1104

; TYPE: DNA

ORGANISM: Ptilosarcus gurneyi
FEATURE:
NAME/KEY: CDS
LOCATION: (34)...(747)
OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

Query Match 17.6%; Score 119; DB 4; Length 1104;
Best Local Similarity 49.4%; Pred. No. 9.2e-15;
Matches 308; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

QY 23 TCACGAGTTCATGCGCTTCAGGTCGCGATGAGGGCACCGTGAACGGCCACGAGTTGCG 82
DB 65 TGAAGAGATTATGTCGGCAAGCTAGCGTTGAAGGAATCGTGAACAATCACGTTTTTT 124
QY 83 AGATCAGGGCGAGGCGGAGGCGCGCCCTACGAGGGCCACAAACACCGTGAAGTTGAAGG 142
DB 125 CCAATGAAGGATTGGAAGGCAATGATATTATTGGAACCAATTGATGCAATCCGG 184
QY 143 TGACCAAGGGCGGCGCCCTGCGCTGCGCTGGGAATCTGTCGCCCCAGTTCCAGTAGC 202
DB 185 TTACAAAGGAGGTCGCTTGCCTTGCCTTTCGATATTGTTCCATAGCTTTTCCAATACG 244
QY 203 GCTCCAGAGTGTACGTGAAGCACCGCGGACATCCCGACATCCCGACATCAAGAGCTGCTCTCC 262
DB 245 GGAATCGCACTTTCAGGAATACCCAGACGACATTCGGGACATCTTTGTTCAATCATTC 304
QY 263 CCGAGGCTTCAAGTGGGAGCGGTGATGAATCTTCAGAGACGCGCGGCTGCGGCGTGA 322
DB 305 CGGCTGGATTTTCTACGAAAGAAATCTACGCTTTGAAGATGGCGCAATTGTTGACATTC 364
QY 323 CCGAGACTCTCCCTGCGAGGAGGCTGCTTCTATCTACAGAGTGAAGTTCAATCGGCTGA 382
DB 365 GTTCAGATATAAGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGCAACG 424
QY 383 ACTTCCCTCCGACGCGCCGCTGATGACAGAAAGACCATGGGCTGGAGGCTCCACCG 442
DB 425 GTTCCCTAGTAGAAGACCGCTGATGACAAAGCCATCTCGCATGGAGCCATCGTTTG 484
QY 443 AGGCGCTGATACCCCGCGAGCGGCTGCTGAAGGGGAGATCCACAGAGCCCTGAAGCTGA 502
DB 485 AGGTGCTTACATGAACAGCGGCTTCTGCTGGGCGAAGTAGATCTCGTTTCAAACTCG 544
QY 503 AGGAGCGGCGCACTACCTGTTGAGTTCAAGTCCATCTACATGGCCAGAGACCGCTGC 562
DB 545 AGTCAGGAACATTAATCTGTCACATGAAACGTTTACAGATCCAAAGGTGGAGTGA 604
QY 563 AGTGCCTGCTACTACTACTGAGACCAAGCTGGACATCACTCCCAACAGGAGTACT 622
DB 605 AAGAAATTCGGGAATATCACTTTATCCATCATGCTCTGGAGAAACCTACGTGGAAGAG 664
QY 623 ACACCATCGTGGAGCAGTACGAG 645
DB 665 GAAGCTTCGGAACACACGAG 687

RESULT 8

US-09-277-716-31
Sequence 31, Application US/09277716A
Patent No. 6232107
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Szent-Gyorgyi, Christopher
APPLICANT: PROLUME, LTD.
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
CURRENT APPLICATION NUMBER: US/09/277,716A
CURRENT FILING DATE: 1999-03-26
EARLIER APPLICATION NUMBER: 60/102,939
EARLIER FILING DATE: 1998-10-01
EARLIER APPLICATION NUMBER: 60/089,367
EARLIER FILING DATE: 1998-06-15
EARLIER APPLICATION NUMBER: 60/079,624
EARLIER FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 31
LENGTH: 1279
TYPE: DNA
ORGANISM: Ptilosarcus gurneyi
FEATURE:
NAME/KEY: CDS
LOCATION: (7)...(720)
FEATURE:
OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-277-716-31

Query Match 17.6%; Score 119; DB 3; Length 1279;
Best Local Similarity 49.4%; Pred. No. 9.3e-15;
Matches 308; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

QY 23 TCACGAGTTCATGCGCTTCAGGTCGCGATGAGGGCACCGTGAACGGCCACGAGTTGCG 82
DB 38 TGAAGAGATTATGTCGGCAAGCTAGCGTTGAAGGAATCGTGAACAATCACGTTTTTT 97
QY 83 AGATCAGGGCGAGGCGGAGGCGCGCCCTACGAGGGCCACAAACCGTGAAGTTGAAGG 142
DB 98 CCATGGAAGGATTGGAAGGCAATGATATTATTGGAACCAATTGATGCAATCCGG 157
QY 143 TGACCAAGGGCGGCGCCCTGCGCTTGCCTGGGACATCTGTCGCCCGTCCAGTACG 202
DB 158 TTACAAAGGAGGTCGCTTGCCTTTCGATATTGTTCCATAGCTTTTCCAATACG 217
QY 203 GCTCCAGAGTGTACGTGAAGCACCGCGGACATCCCGACATCCCGACTCAAGAGCTGCTCTCC 262
DB 218 GGAATCGCACTTTCAGGAATACCCAGACGACATTCGGGACTACTTTGTTCAATCATTC 277
QY 263 CCGAGGCTTCAAGTGGGAGCGGTGATGAATCTGAGGAGCGGGCGTGGCGGCGTGA 322
DB 278 CGGCTGGAATTTTCTACGAAGAAATCTACGCTTTGAAGATGGCGCCATTTTGACATTC 337
QY 323 CCGAGGACTCTCCCTCGAGGACCGCTGCTTCTATCTACAAGGTGAAGTTCAATCGGCTGA 382
DB 338 GTTCAGATATAAGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGCAACG 397
QY 383 ACTTCCCTCGAGCGGCGCGCTGATGACAGAAAGACCATGGGCTGGAGGCTCCACCG 442
DB 398 GTTTCCTCTAGTAACGAGCCGCTGATGCAAAAGCCATCTCGGATGAGGCGCATCGTTTG 457
QY 443 AGCGCTGTATCCCGCGAGCGGCTGCTGAAGGCGGAGATCCCAAGGCCCTGAAGCTGA 502
DB 458 AGGTGCTTACATGAACAGCGGCTTCTGTTGGGCGAAGTAGATCTCGTTTACAACCTCG 517
QY 503 AGGAGCGGCGCACTACTGCTGGTGAAGTTCAAGTCCATCTACATGGCCCAAGAGCCGTCG 562
DB 518 AGTCAGGAACATATTACTCGTCCACATGAAGGCTTTTACAGATCCAAAGGTGGAGTGA 577
QY 563 AGCTGCCCGCTACTACTACTGAGACCAAGCTGGACATCACTCCCAACAGGAGTACT 622
DB 578 AAGAATTCGCGAATATCACTTTATCCATCATCTGTTGGAAGAAACCTACGTGGAAGAG 637
QY 623 ACACCATCGTGGAGCAGTACGAG 645
DB 638 GAAGCTTCGGAACACACGAG 660

RESULT 9

US-09-609-161B-31
Sequence 31, Application US/09609161B
Patent No. 6436682
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Szent-Gyorgyi, Christopher
APPLICANT: PROLUME, LTD.
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIC
TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

```
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-609-161B-31

Query Match      17.6%; Score 119; DB 4; Length 1279;
Best Local Similarity 49.4%; Pred. No. 9.3e-15; Indels 0; Gaps 0;
Matches 308; Conservative 0; Mismatches 315;

Qy 23 TCACCGAGTTCATCGGCTTCAAGTGGCGATGAGGCGCACCGTGAACGGCCACGAGTTGCG 82
Db 38 TGAAGAAGATATGTGCGCAAAAGTAGCGTTGAAGAAATCGTGAACAATCACGTTTTTT 97

Qy 83 AGATCGAGGGCGGGCGAGGGCGGCCCTTCAGAGGGCCACACACCGTGAAGTTGAAG 142
Db 98 CCATGGAAGGATTTGGAAGAGGCAATGATATATTTGGAACCAATGATGCAAAATCCGGG 157

Qy 143 TGACCAAGGGCGGGCGGCCCTTCGCTGGGACATCTGTCGCCCGCCAGTTTCCATAGC 217
Db 158 TTACAAAGGAGGTCGTTGCCATTCGTTTCGACATTTGTTCCATAGCTTCCATAGC 217

Qy 203 GCTCAAGAGTGTAGTGAAGCACCCCGCGACATCCCGGACTACAGAAAGCTGTCTCTCC 262
Db 218 GGAATCGCACTTTCACGAAATATCCAGACGACATTCGGGACTACTTTGTTCAATCATTC 277

Qy 263 CCGAGGCTTCAAGTGGGCGGCTGATGAATTCGAGGACGGCGCGTGGCGACCGTGA 322
Db 278 CGGCTGATTTTCTAGAAAGAAATCTACGCTTTGAAGTGGCGGCTTTGACATTC 337

Qy 323 CCAGAGACTCCTCCCTCAGGAGGGTGTCTTCACTAAGTGAAGTCAATCGGGTGA 382
Db 338 GTTCAGATATAAGTTTGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAAG 397

Qy 383 ACTTCCCTCCGAGGCGCGCTGATCAGAAAGAACCATGGCTGGGAGGCTCCACCG 442
Db 398 GTTTCCTCTAGTACGAGACCGTGTATGCAAAAGCCATCTCGGATGGAGCCATCTTTG 457

Qy 443 AGCGCTGTATCCCGCGAGCGGGTGTGTAAGGGCGAGATCCCAAGGCGCTGAAGCTGA 502
Db 458 AGGTGCTCATATGAACAGCGCGGTTCTGGTGGCGAAGTAGATCTCGTTTACAACTCG 517

Qy 503 AGAGCGGGCCACTACTGTGGAGTTCAAGTCAATCTACATGGCCAGAGCCCGTGC 562
Db 518 AGTCAGGAATCTTACTCTGCGCATGAAACGTTTTTACAGATCCAAAGGTGGAGTGA 577

Qy 563 AGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACTCCCAACAGGAGCT 622
Db 578 AAGAATTCGCGAATATCACTTTATCATCATCGTCTGGAGAAACCTACGTGGAAGAG 637

Qy 623 ACACCATCGTGGAGAGTAGAG 645
Db 638 GAAGCTTCGTGGAACAACAGAG 660
```

RESULT 10

```
US-09-459-956-4
; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-4

Query Match      16.3%; Score 110.4; DB 4; Length 696;
Best Local Similarity 52.5%; Pred. No. 4.1e-13;
Matches 296; Conservative 0; Mismatches 256; Indels 12; Gaps 2;

Qy 1 ATGGCTCTCTCCGAGAACGTTCATCCGAGTTTCAGGTTCAAGTGGCGATCGAGGGC 60
Db 1 ATGGCTCATTCAAAGCAGCGTCTAAAGAGAGAAATGAATGAATACCAATGGAAGGG 60

Qy 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGCGAGGGCGAGGGCCGCCCTACGAGGGC 120
Db 61 TCGCTCAACGAGACATAAATTTGATCAGCGGCGAGGATTCGATATCCGTTCAAGGG 120

Qy 121 CACACACCGTGAAGTTGAAGTGCACCAAGGGCGGCCCTGCGCTTCGCTGGGACATC 180
Db 121 AAACAGACTATTAATCTGTGTGTGATCGAAGGGGACANTTCCATTTTCGAAGACATA 180

Qy 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 TTGTCAAGTGGCTTTAAGTACGGAGACAGGATTTTCACTGAATATCTCAAGACATAGTA 240

Qy 241 GACTACAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACTTCGAG 300
Db 241 GACTATTTCAAGAACTCGTGTCTGTGGATATACATGGGGGAGGTCTTTTCTTTTGG 300

Qy 301 GACGCGCGGTGGCGGACCGTGACCCGAGACTCTCTCCCTG-----CAGGACGGCTGCTTC 354
Db 301 GATGGAGCAGTCTCATATGCAATGTAGATATAACAGTGTGTCAAAGAAACTGCAAT 360

Qy 355 ATCTAAGAGTGAAGTTTCATCGGCGTGAACCTTCCCTTCGACGCGCCCGGTGATGAGAAG 414
Db 361 TATCATGAAGACATATTTAATGGAATGAATTTTCTCTGATGAGACCTGTGATGAAAAAG 420

Qy 415 AAGACCATGGCTGGGAGGCTCCACCGAG-----CGCCTGTACCCCGCGAGCGGTG 468
Db 421 ATGACACTAATCTGGAGAGCATCTGCGAGAGATCATGCCAGTACCTAGACGGGATA 480

Qy 469 CTGAAGGCGAGATTCACAAAGGCCCTGAAAGCTGAAGACGCGGCGCACTACTCTGTGGAG 528
Db 481 CTGAAGGGAGTGTCTCCATGTACCTCTTCTGAAGGATGTTGGGGCTTACCGGTGCGAG 540

Qy 529 TTCAAGTCCATCTCATGCGCAAG 552
Db 541 TTCGACACAGTTTACAAAGCAAG 564

RESULT 11
US-09-214-909-1
; Sequence 1, Application US/09214909
```


Patent No. 6486382
GENERAL INFORMATION:
APPLICANT: GORDON-KAMM, WILLIAM
APPLICANT: PIERCE, DOROTHY A.
APPLICANT: BOWEN, BENJAMIN
APPLICANT: BIDNEY, DENNIS
APPLICANT: ROSS, MARGIT
APPLICANT: SCHELONGE, CHRISTOPHER
APPLICANT: MILLER, MICHAEL D.
APPLICANT: SANDAHL, GARY
APPLICANT: WANG, LIJUAN
TITLE OF INVENTION: USE OF THE GREEN FLUORESCENT PROTEIN AS A SCREENABLE MARKER FOR PLANT TRANSFORMATION
FILE REFERENCE: 033229/0682
CURRENT APPLICATION NUMBER: US/09/214,909
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US97/07688
PRIOR FILING DATE: 1997-05-01
PRIOR APPLICATION NUMBER: 60/016,345
PRIOR FILING DATE: 1996-05-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 717
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
OTHER INFORMATION: encoding GFP
NAME/KEY: CDS
LOCATION: (1)..(714)
US-09-214-909-1

Query Match 14.8%; Score 100.2; DB 4; Length 717;
Best Local Similarity 50.1%; Pred. No. 3.7e-11;
Matches 336; Conservative 0; Mismatches 323; Indels 12; Gaps 3;

QY 1 ATGGCTCTCTCGAGAACGTCATCACGAGTTTCATCGGCTTCAAGTGGCATGAGGGC 60
Db 1 ATGTCAAGGGCGAGAGGCTCTTACCGGGCTGTGCTCCATCTCTGTGAGTCTGACGGC 60
QY 61 ACCTGTAAGCGGCACAGTTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTACGAGGGC 120
Db 61 GAGTGAAGCGGCACAGTTTCGAGTGTGCGGCGAGGGCGAGGGCGAGGGCGCCCTACGAGC 120
QY 121 CACAACACCGTGAAAGTTGAAGTGACCAAGGGCGGGCCCTGCTTCCGCTGGGACATC 180
Db 121 AAGTCAACCTCAAGTTTCATCTGACCA---CCGGCAAGCTCCCGCTGCTGGCCACG 177
QY 181 CTGTCCCCCAGTTCAGTACGGCTCCAAGTGTGAGTGAAGCAGCCCGCGACAT- 236
Db 178 CTGTGACCACTTCTCTACGCGTGCAGTGTCTTCCAGGTACCCCGACCATGAAG 237
QY 237 -CCCGCACTACAGAAAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAAC 294
Db 238 CAGCAACACTTTCAGTACGATGCGCCGAGGGCTACGTGAGGAGGAGGACCATCTTC 297
QY 295 TTCAGGACCGCGCGTGGCGACCGTGACCCAGACTTCTCTCCCTGAGGAGCGCTGCTTC 354
Db 298 TTCAGGACCGCGCAACTACAGACCAAGCGCGCGAGGTGAAGTTCGAAGCGCACCCCTC 357
QY 355 ATCTCAAGGTGAAGTTTCATCGGCTGAATCTTCCCTCCGACGCGCCCGTGAATGCAAG 414
Db 358 GTGAACAGGATTTAGCTTAAGGGATTCGATTTCAAGGAGGAGCGGACATCTTCGCGCCAC 417
QY 415 AAGACCATGGGCTGGGAGGCTTCCACGAGCGCTGTACCCCGGAGCGCGTGTGAAG 474
Db 418 AAGCTCAGTACAACTCAACTCCCAAA---CGTGTACATCATGCGGAGGAGGAGGAG 474
QY 475 GCGAGATCCCAAGGGCCTGAAGTGAAGGACGCGCGGCGCTACCTGTGGAGTTCAAG 534
Db 475 AACGGCATCAAGGTGAATTCAGATCAGGCACAAATCAGGAGCGCTCAGTGCAGCTC 534

QY 535 TCCATCTTACATGGCCCAAGAGCCGTCAGCTCCCGGCTACTACTACGTGGACACCAAG 594
Db 535 GCTGACCACTACCAGCAGAACACCCCATCGGAGCGGCCCGCTGTCTTCCCGCAAC 594
QY 595 CTGACATACCTCCACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAG 654
Db 595 CACTACCTTCCACACAGTCCGCCCTCTCCAGGACCCCAACGAGAGGAGCCACATG 654
QY 655 GGCGGCCACCA 665
Db 655 GTGCTCTCTGA 665

RESULT 12
US-09-459-956-3
Sequence 3, Application US/09459956
Patent No. 6342379
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Gonzalez, Ili, Jesus E.
TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
TITLE OF INVENTION: OPTICAL METHODS
FILE REFERENCE: REGEN1290-4
CURRENT APPLICATION NUMBER: US/09/459,956
CURRENT FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: 08/765,860
PRIOR FILING DATE: 1999-05-08
PRIOR APPLICATION NUMBER: 08/481,977
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: PCT/US96/09652
PRIOR FILING DATE: 1996-06-06
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 696
TYPE: DNA
ORGANISM: Zoanthus sp
US-09-459-956-3

Query Match 14.6%; Score 99.2; DB 4; Length 696;
Best Local Similarity 51.2%; Pred. No. 5.8e-11;
Matches 289; Conservative 0; Mismatches 263; Indels 12; Gaps 2;

QY 1 ATGGCTCTCTCGAGAACGTCATCACGAGTTTCATCGGCTTCAAGGTGCGCATGAGGGC 60
Db 1 ATGGTCAAGTCAAGCAGGCTTACAAAAGAAATGACATGAAATACCGTATGGAAGG 60
QY 61 ACCGTGAAGCGCACGAGTTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTACGAGGGC 120
Db 61 TCGCTGATGACATATAATTTGTGATCACGGGAGAGGGCATTTGGATATCCGTTCAAGGG 120
QY 121 CACACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTGCTTCCGCTGGACATC 180
Db 121 AAACAGGCTATTATCTGTGTGTGTCGAAGTGAACATTTGCCATTTGCCGAGACATA 180
QY 181 CTGTCCCCCAGTTCAGTACGGTTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 240
Db 181 TTGTGAGTCTTAACTTAACTAGGAAACAGGGTTTCACTGAATATCTTCAAGACATAGT 240
QY 241 GACTACAGAGTCTTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAATTCGAG 300
Db 241 GACTATTTCAAGAACTCGTGTCTGATATATCATGGACAGGTCTTTCTCTTTGAG 300
QY 301 GAGCGGGCGGTGGCAGCGGTGACCCAGGACTCTTCCCTG-----CAGGACGGTGTTC 354
Db 301 GATGAGCAGTTTGCATATGTAATGACAGTGAATGAGTGAAGAACTGCATG 360
QY 355 ATCTCAAGGTGAAGTTTCATCGGCTGAATCTTCCCTCGAGCGGCCGTGTGAGAG 414
Db 361 TATCATGATCCAAATTTTATGAGTGAATTTTCTGTGTGATGAGACCTGTGTGATGAAAG 420
QY 415 AAGACCATGGGCTGGAGGCGCTCCACCGAG-----CGCTGTACCCCGGAGCGCTG 468

Db 421 ATGACAGATAACTGGAGGCCATCTCTGCGAGAAGATCATACCAGTACCTAAGACGGGGATA 480
Qy 469 CTGAAGGGCGAGATCCACAAAGCCCTGAAGCTGAAGGACGGCGCCACTACCTGGTGAG 528
Db 481 TTGAAGGGGATGCTCTCATGTACTCTCTCTGAAGGATGTGGCGGCTTACGGTGCCAA 540
Qy 529 TTCAAGTCCATCATATGGCCAAAG 552
Db 541 TTCGACACAGTTTACAAAGCAAG 564

RESULT 13
US-09-486-241-31
; Sequence 31, Application US/09486241
; Patent No. 6472184
; GENERAL INFORMATION:
; APPLICANT: Hegemann, Peter
; TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEIC ACID
; FILE REFERENCE: 3910/06706
; CURRENT APPLICATION NUMBER: US/09/486,241
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/EP98/05219
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: DE19736591.4
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified gene from Aquorea victorea
US-09-486-241-31

Query Match 14.6%; Score 99; DB 4; Length 717;
Best Local Similarity 51.0%; Pred. No. 5.4e-11;
Matches 327; Conservative 0; Mismatches 290; Indels 24; Gaps 3;

Qy 1 ATGCGCTCTCCGAGACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGCATGGAGGC 60
Db 1 ATGTCCAAAGGCGGAGGAGCTGTTACCGGTGTGTCTCCATCTCTGTGGAGCTGGACGGC 60

Qy 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGGCGCCCTACGAGGC 120
Db 61 GACGTGAACGCCACAAAGTTCTCGGTCTCGGCGAGGTTGAGGTGACGCGCACCTACGGC 120

Qy 121 CACAACACCGTGAAGTTGAAGTGAACCAAGGCGGCGCCCTCGCTTCCGCTTGGGACATC 180
Db 121 AAGTGACCTGAACTTCACTGTGACCA---CCGGCAAGCTGCGCCGTGCGCTGGCCACC 177

Qy 181 CTGTCCCGCCAGTTCAGTACGCTCCAGGTGTAGTGAAGCACCCCGCCGACAT---- 236
Db 178 CTGTGTACCACTTCACTACGCTGTGAGTGTCTTCGCGTACCCCGACACATGAAG 237

Qy 237 --CCCCGACTACAAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAAC 294
Db 238 CAGCAGCACTTCTCAAGTCCGCCATGCGGAGGGTACGTGAGAGCGGACCATCTTC 297

Qy 295 TTGAGGACCGGCGGTGGGCGACCGTGAACCGAGTCTCTCCCTGAGGACGGCTGCTTC 354
Db 298 TTCAAGGACACGCGCACTACAAGACCCGCGCGAGGTCAAGTTCAGGCGGACACCCCTG 357

Qy 355 ATCTACAAGTGAAGTTCACTGCGGTGAATCTCCCTCCGAGCGCCCGTGTGATGAGAG 414
Db 358 GTGACCGCATCGAGTGAAGGSCATCGACTTCAAGAGGACGGCAACATCTTGGCCAC 417

Qy 415 AAGACCATGGCTGGGAGGCTCCACCGAGCGCTGTATCCCCGCGACCGCGTGTGAGAG 474
Db 418 AAGCTGAGTACAACTACAACTCCCAACAGCTGTATCATATGCGCCGACAGAGAGAAC 477

Qy 475 GCGGAGATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCACTACCTGGTGGAGTT---- 530

Db 478 GGCATCAAGGTGAATCTCAAGATCCGCCACAACTCGAGGACGGCTCGTGCAGCTGGCC 537
Qy 531 -----CAAGTCCATCTAGTGGCCAAAGACCCGTCGAGCTGCCCGGCTACTAC 579
Db 538 GACCACTACAGCAACACCCCATCGCGCATGGCCCGCTGCTGCTCCCGACAACAC 597
Qy 580 TACGTGGACACCAAGCTGCAATCACCTCCCAACACGAGGA 620
Db 598 TACCTGTCCACCCAGTCCGCCCTGTCCAAGGACCCCAAGA 638

RESULT 14
US-09-609-161B-15
; Sequence 15, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUNE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-15

Query Match 14.5%; Score 98.4; DB 4; Length 1079;
Best Local Similarity 50.4%; Pred. No. 8.5e-11;
Matches 240; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

Qy 28 GAGTTCATGCGCTTCAAGTGGCGCATCGAGGCGCACCGTGAAGGTCGAGATC 87
Db 295 GAAGTAATGTCGTATAAGTAATCTGGAAGGAATTTGTAACACCACTGTTTACAACT 354

Qy 88 GAGGCGGAGGCGGCGCGCCCTACGAGGCGCACCAACCGGTGAAGTGTGAAGGTGACC 147
Db 355 GAGGTTTCGGCAAGGGAATATTTTATTCGCAATCACTGGTTTCAGATTCGTGTCAAG 414

Qy 148 AAGGCGGCGCGCTGCGCTGCGGACATCTGTCCCGCCAGTTCGAGTACGCTCC 207
Db 415 AAGGCGGCGCGCTGCGCTTGTGATTTGATTTGTCACCACTTTCATTAATGGAAC 474

Qy 208 AAGGTGACGTGAAGACACCGCGCATCCCGACTACAAGAGCTGTCTTCCCGGAG 267
Db 475 CGTACTTTCAGAAATATCCGAATGATATATCAGATTTATTTATACATCATTTCCAGCA 534

Qy 268 GGCTTCAAGTGGGAGCGGTGATGAATTCGAGACCGCGCGGTGGGACCGTGACCCAG 327
Db 535 GGATTTATGATGAACGAACATTAACGTTACGAGAGATGGCGGACTTGTGAAATTCGTCA 594

Qy 328 GACTCCCTCCGACGAGGCGCTGCTTCATCTACAAGTGAAGTTCATTCGCGCGTGAATTC 387
Db 595 GATATAAATTTAATAGAGCAAGTTCGTCTACAGAGTGAATACAAGGTAGTACTTC 654

388	CCCTCCGACGCCCCCGTGTATGCGAAGAACCATCGGCTGGGAGCGCTCCACCGAGCGC	447
QOY		
655	CCAGATGATGTCGCCCGTCATGCAAGAAGCTATCTTAGGAATAGAGCCTTCATTTGAAGCC	714
DBD		
448	CTGTACCCCCCGCAGCGCGTCTCAAGCGGCGAGATCCACAAGGCCCTCAAGCTGAA	503
QOY		
715	ATGTACATGAATATATGCGCTTGGTCCGCGAGTAACTCTTGCTATAACTTAA	770
DBD		

```

RESULT 15
US-09-626-581D-64
; Sequence 64, Application US/09626581D
; Patent NO. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Functions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-626-581D-64

```

Query Match	14.5%	Score 98.4;	DB 4;	Length 1079;
Best Local Similarity	50.4%;	Pred. No. 8.5e-111;		
Matches 240;	Conservative 0;	Mismatches 236;	Indels 0;	Gaps 0;
28	QY	GAGTTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTCGAACGGGCCACGAGTTCGAATC 87		
295	Db	GAAGTAATGTCGTATAAAGTAAATCTGGAAGGAATTGTAACAACCATGTTTTACAATG 354		
88	QY	GAGGGCGAGGGCGAGGGCCGCCCTTCAGAGGGGCCAACACCGCTGAAGTTGAGGTGACC 147		
355	Db	GAGGGTTTCGGGCCAAGGGGAATATTTATTTCGGCAATCAACTCGTTCAGATTGCTGTCAAG 414		
148	QY	AAGGGCGGCCCGCTTCGCTTGGCATCTCTGTCGCCCCAGTTCACGATACGGCTCC 207		
415	Db	AAAGGGGCCCACTGCCCTTTTGCAATTGATTTGTGTCACCAAGCTTTTCAATATGGCAAC 474		
208	QY	AAGGTGTACGTGAAGCAACCGCCGGCAGATCCCGCATCAAGAAGGTGTCTCTTCCCGAG 267		
475	Db	CGTACTTTTCAGGAATATCCGAATGATATACAGATTATTTTATCAATCATATTTCCAGCA 534		
268	QY	GGCTTCAAGTGGGAGCGGTGATGAATCTCGAGGAGCGGGCGTGGCGACCGTGAACCCAG 327		
535	Db	GGATTTATGTATGAACGAAACATTACGTTACAGAGATGGCGGACTTGTGGAAATTCGTTC 594		
328	QY	GACTCTCTCCCTGCAGGACGGCTGCTTCATCTCAAGGTGAAGTTCATCGGCGTGAACCTTC 387		
595	Db	GATATAAAATTTAATAGAAGCAACAGTTCGTCTACAGAGTGGAAATACAAAGGTAGTAACCTTC 654		
388	QY	CCCTCCGACGCGCCGCTGATGCGAAGAAGACCATGGCTGGGAGCGCTCCACCGAGCGC 447		
655	Db	CCAGATGATGTCGCCGTATGCGAGAGACTATCTTAGGAATAGAGCCTTCATTGGAGCC 714		
448	QY	CTGTACCCCCCGACGGCTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAA 503		
715	Db	ATGTACATGAATAATGGCTCTTGGTCGGCGGAAGTAATCTTGTCTATAAACTAAA 770		

Blank Sheet U.S. PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 14:04:25 ; Search time 492.957 Seconds
(without alignments)

6743.643 Million cell updates/sec

Title: US-10-081-864-14

Perfect score: 678

Sequence: 1 atggctctctcgagaacgt.....gccaccactgttctgttaa 678

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	678	100.0	678	14	US-10-006-922-36
2	678	100.0	678	15	US-10-081-864-14
3	676.4	99.8	678	15	US-10-315-920-5
4	672.2	99.1	1050	14	US-10-060-857-7
5	670.2	98.8	675	14	US-10-006-922-38
6	670.2	98.8	675	15	US-10-081-864-13
7	668.4	98.6	678	15	US-10-315-920-3
8	665.2	98.1	678	15	US-10-081-864-7
9	665.2	98.1	678	15	US-10-315-920-1
10	661.6	97.6	681	15	US-10-121-258-3
11	661.6	97.6	681	15	US-10-121-258-23
12	661.6	97.6	4632	15	US-10-161-403-29
13	661.6	97.6	4632	17	US-10-433-640-16
14	661.6	97.6	6984	14	US-10-001-189-45

15	660	97.3	681	14	US-10-006-922-35	Sequence 35, Appl
16	660	97.3	681	14	US-10-006-922-37	Sequence 37, Appl
17	659.6	97.3	1638	15	US-10-214-932-51	Sequence 51, Appl
18	659.6	97.3	1647	15	US-10-214-932-75	Sequence 75, Appl
19	655.2	96.6	723	15	US-10-152-296-1	Sequence 1, Appl
20	655.2	96.6	723	17	US-10-739-656-1	Sequence 1, Appl
21	649.8	95.8	678	15	US-10-121-258-5	Sequence 5, Appl
22	629.6	92.9	681	15	US-10-121-258-7	Sequence 7, Appl
23	568.6	83.9	675	17	US-10-121-258-15	Sequence 15, Appl
24	568.6	83.9	678	15	US-10-121-258-9	Sequence 9, Appl
25	566.4	82.1	678	14	US-10-006-922-43	Sequence 43, Appl
26	556.4	82.1	678	15	US-10-081-864-11	Sequence 11, Appl
27	470.6	69.4	549	17	US-10-724-178-1040	Sequence 1040, Ap
28	459	67.7	561	17	US-10-724-178-1032	Sequence 1032, Ap
29	438.4	64.7	504	17	US-10-724-178-1036	Sequence 1036, Ap
30	438	64.6	678	9	US-09-967-772-6	Sequence 6, Appl
31	438	64.6	678	13	US-10-132-067-3	Sequence 3, Appl
32	438	64.6	678	14	US-10-006-922-11	Sequence 11, Appl
33	438	64.6	678	15	US-10-121-258-2	Sequence 2, Appl
34	438	64.6	678	16	US-10-335-517-6	Sequence 6, Appl
35	438	64.6	678	16	US-10-334-288-6	Sequence 6, Appl
36	438	64.6	859	9	US-09-999-745-66	Sequence 66, Appl
37	438	64.6	859	10	US-09-866-538-11	Sequence 11, Appl
38	438	64.6	859	10	US-09-794-308-11	Sequence 11, Appl
39	438	64.6	859	10	US-09-865-291-11	Sequence 11, Appl
40	438	64.6	859	17	US-10-433-640-12	Sequence 12, Appl
41	437	64.5	3311	10	US-09-797-496B-3	Sequence 3, Appl
42	432	63.7	666	17	US-10-332-733-22	Sequence 22, Appl
43	426	62.8	711	17	US-10-314-936-1	Sequence 1, Appl
44	426	62.8	711	17	US-10-314-936-3	Sequence 3, Appl
45	422	62.2	898	14	US-10-006-922-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1

US-10-006-922-36
; Sequence 36, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-36

Query Match 100.0%; Score 678; DB 14; Length 678;
Best Local Similarity 100.0%; Pred. No. 4.4e-163;

Matches	678;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ATFGGCTCTCCGAGAA	CGTCATCAC	CGAGTTCA	TCGCTTCA	AGTGCGCATG	AGGGC	60	
Db	1	ATFGGCTCTCTCCGAGAA	CGTCATCAC	CGAGTTCA	TCGCTTCA	AGTGCGCATG	AGGGC	60	
QY	61	ACCGTGAACGCCAC	CGAGTT	CGAGAT	CGAGGCGAGG	CGAGGCGCGCCCT	TACGAGGC	120	
Db	61	ACCGTGAACGCCAC	CGAGTT	CGAGAT	CGAGGCGAGG	CGAGGCGCGCCCT	TACGAGGC	120	
QY	121	CACAAACCGGTGA	AGTTGA	AGGTGAC	CAACAGGCGCGCCCT	TGCTTGCCTTGGGACATC	180		
Db	121	CACAAACCGGTGA	AGTTGA	AGGTGAC	CAACAGGCGCGCCCT	TGCTTGCCTTGGGACATC	180		
QY	181	CTGTCCCGCCAGT	TCCAGTAC	GGCTTCA	AGGTGTA	AGTGAAGCACCCGCGGACATCCCC	240		
Db	181	CTGTCCCGCCAGT	TCCAGTAC	GGCTTCA	AGGTGTA	AGTGAAGCACCCGCGGACATCCCC	240		
QY	241	GACTACAAAGAAG	CTGTCTT	CCCCGAGGG	CTTCAAGTGGGAGCGGT	GATGAACCTTCGAG	300		
Db	241	GACTACAAAGAAG	CTGTCTT	CCCCGAGGG	CTTCAAGTGGGAGCGGT	GATGAACCTTCGAG	300		
QY	301	GACGCGCGGTGG	GCACCGGTGA	CCAGGACTCT	CCCTGCAGAGACGGCTGCTTCA	TCTATCTAC	360		
Db	301	GACGCGCGGTGG	GCACCGGTGA	CCAGGACTCT	CCCTGCAGAGACGGCTGCTTCA	TCTATCTAC	360		
QY	361	AAGGTGAAGTTCA	TGCGGTGA	CTTCCCTCCGAGCG	CCCGCTGATGCAGAAAGAAGACC	420			
Db	361	AAGGTGAAGTTCA	TGCGGTGA	CTTCCCTCCGAGCG	CCCGCTGATGCAGAAAGAAGACC	420			
QY	421	ATFGGCTGGGAGG	CGCTTCA	CCGAGCGCTGT	TACCCCGCGACGGGCTGTAAGAGGGCGAG	480			
Db	421	ATFGGCTGGGAGG	CGCTTCA	CCGAGCGCTGT	TACCCCGCGACGGGCTGTAAGAGGGCGAG	480			
QY	481	ATCCACAAAGGCC	TGAAGCTGA	AGGACGGGCGGCACT	TACCTGCTGGAGTTCAAGTCCATC	540			
Db	481	ATCCACAAAGGCC	TGAAGCTGA	AGGACGGGCGGCACT	TACCTGCTGGAGTTCAAGTCCATC	540			
QY	541	TACATGGCCAAAGA	AGCCCGTGC	AGCTGTC	CCGGCTACTACTAGTGGACACCAAGCTGGAC	600			
Db	541	TACATGGCCAAAGA	AGCCCGTGC	AGCTGTC	CCGGCTACTACTAGTGGACACCAAGCTGGAC	600			
QY	601	ATCACCTCCACAC	ACGAGGACTAC	ACCATGCTGGAGCACT	TACAGCGCACCCGAGGGCGC	660			
Db	601	ATCACCTCCACAC	ACGAGGACTAC	ACCATGCTGGAGCACT	TACAGCGCACCCGAGGGCGC	660			
QY	661	CACCACCTGTTCT	CTGTAA	678					
Db	661	CACCACCTGTTCT	CTGTAA	678					

RESULT 2
 US-10-081-864-14
 ; Sequence 14, Application US/10081864
 ; Publication No. US2003002287A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey
 ; APPLICANT: Lukyanov, Konstantin
 ; APPLICANT: Yanushevich, Yuriy
 ; APPLICANT: Savitsky, Alexandr
 ; APPLICANT: Fradkov, Arcady
 ; TITLE OF INVENTION: No. US2003002287A1 Aggregating Fluorescent Proteins and
 ; TITLE OF INVENTION: Methods for Using the Same
 ; FILE REFERENCE: CLON-067
 ; CURRENT APPLICATION NUMBER: US/10/081,864
 ; CURRENT FILING DATE: 2002-06-19
 ; PRIOR APPLICATION NUMBER: 10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 60/270,983
 ; PRIOR FILING DATE: 2001-02-21
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14

```

; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-14

Query Watch
Best Local Similarity 100.0%; Score 678; DB 15; Length 678;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 ATGGGCTCTCCGAGAAGCTCATCACCGAGTTTCATCGCTTCAAGTGGCATGAGGCG 60
Db 1 ATGGGCTCTCCGAGAAGCTCATCACCGAGTTTCATCGCTTCAAGTGGCATGAGGCG 60

Qy 61 ACCGTGAAGGCGCCAGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTACGAGGCG 120
Db 61 ACCGTGAAGGCGCCAGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTACGAGGCG 120

Qy 121 CACACACCGTGAAAGTTCGAAAGTTCACCAAGGGGGGCGCCCTGCGCTTCGCCCTGGGACATC 180
Db 121 CACACACCGTGAAAGTTCGAAAGTTCACCAAGGGGGGCGCCCTGCGCTTCGCCCTGGGACATC 180

Qy 181 CTGTCCCCCGAGTTCAGTAGCGGTCCAAGGTGTAGTGAAGCACCCCGGACATCCCC 240
Db 181 CTGTCCCCCGAGTTCAGTAGCGGTCCAAGGTGTAGTGAAGCACCCCGGACATCCCC 240

Qy 241 GACTACAAAGAGCTGTCTTCCCGAGGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAG 300
Db 241 GACTACAAAGAGCTGTCTTCCCGAGGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAG 300

Qy 301 GACGCGCGGTGGCGACCGGTGACCCAGGACTCTCTCCCTGGAGGACGGCTGTCTATCTAC 360
Db 301 GACGCGCGGTGGCGGACCGGTGACCCAGGACTCTCTCCCTGGAGGACGGCTGTCTATCTAC 360

Qy 361 AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGAGCGGCCCGGTGATGACGAAGAAGACC 420
Db 361 AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGAGCGGCCCGGTGATGACGAAGAAGACC 420

Qy 421 ATGGGCTGGGAGGCTTCCACGAGCGGCTGTATACCCCGCGAGCGGCGTGTGAAGGGCGAG 480
Db 421 ATGGGCTGGGAGGCTTCCACGAGCGGCTGTATACCCCGCGAGCGGCGTGTGAAGGGCGAG 480

Qy 481 ATCCACAGGCGCTGAAGCTGAAGGACGCGGGCCACTACTCTGTGGAGTTCAAGTCCATC 540
Db 481 ATCCACAGGCGCTGAAGCTGAAGGACGCGGGCCACTACTCTGTGGAGTTCAAGTCCATC 540

Qy 541 TACATGGCCAAAGAGCCGCTGCAGCTGCCCGGTACTACTATGCTGGACACCAAGCTGGAC 600
Db 541 TACATGGCCAAAGAGCCGCTGCAGCTGCCCGGTACTACTATGCTGGACACCAAGCTGGAC 600

Qy 601 ATCACTCCACACAGGAGGACTACCCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCG 660
Db 601 ATCACTCCACACAGGAGGACTACCCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCG 660

Qy 661 CACCACTGTTTCCTGTAA 678
Db 661 CACCACTGTTTCCTGTAA 678

```

7

RESULT 3
US-10-315-920-5
; Sequence 5, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607

; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-5

Query Match 99.8%; Score 676.4; DB 15; Length 678;
Best Local Similarity 99.9%; Pred. No. 1.1e-162;
Matches 677; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC	60
Db	1	ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC	60
QY	61	ACCGTGAACGGCCACAGTTCGAGATCGAGGGGAGGGGCGGGCGGCCCTACGAGGC	120
Db	61	ACCGTGAACGGCCACAGTTCGAGATCGAGGGGAGGGGCGGGCGGCCCTACGAGGC	120
QY	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTGCCTTGCCTTGGGACATC	180
Db	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTGCCTTGGGACATC	180
QY	181	CTGTCCCTCCAGTTCAGTACGGCTCAAGGTGACGTAAGGACCCCGCCGACATCCCC	240
Db	181	CTGTCCCTCCAGTTCAGTACGGCTCAAGGTGACGTAAGGACCCCGCCGACATCCCC	240
QY	241	GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGGCGGTGATGAACTTCGAG	300
Db	241	GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGGCGGTGATGAACTTCGAG	300
QY	301	GACGGGGGTGGCGACCGTGAAGGACCTCCCTCCGAGGACCGGTGCTTCTATCTAC	360
Db	301	GACGGGGGTGGCGACCGTGAAGGACCTCCCTCCGAGGACCGGTGCTTCTATCTAC	360
QY	361	AGGTGAAGTTCTCCGCTGAATTCCTCCGAGGCGCCCGTGTGAGAGAGAGACC	420
Db	361	AGGTGAAGTTCTCCGCTGAATTCCTCCGAGGCGCCCGTGTGAGAGAGAGACC	420
QY	421	ATGGGCTGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGCGTGTGAAGGGCGAG	480
Db	421	ATGGGCTGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGCGTGTGAAGGGCGAG	480
QY	481	ATCCAAAGCCCTGAAGTGAAGGACGGCGGCACTACCTGTGAGTTCAAGTCCATC	540
Db	481	ATCCAAAGCCCTGAAGTGAAGGACGGCGGCACTACCTGTGAGTTCAAGTCCATC	540
QY	541	TACATGCCCAAGAGCCCGTGCAGCTGCCCGCTACTACTGTGACACCAAGTGGAC	600
Db	541	TACATGCCCAAGAGCCCGTGCAGCTGCCCGCTACTACTGTGACACCAAGTGGAC	600
QY	601	ATCACCTCCCAACAGGAGCTACACATCTGTGGAGCAGTACGAGCGACCGAGGGCGC	660
Db	601	ATCACCTCCCAACAGGAGCTACACATCTGTGGAGCAGTACGAGCGACCGAGGGCGC	660
QY	661	CACCACCTGTTCTCTGTA 678	
Db	661	CACCACCTGTTCTCTGTA 678	

RESULT 4
US-10-060-857-7
; Sequence 7, Application US/10060857
; Publication No. US20020132318A1
; GENERAL INFORMATION:
; APPLICANT: Anticancer, Inc.
; APPLICANT: Zhao, Ming

; APPLICANT: Jiang, Ping
; APPLICANT: Xu, Mingxu
; APPLICANT: Yang, Meng
; TITLE OF INVENTION: IMPROVED FLUORESCENT PROTEINS
; FILE REFERENCE: 31276-20032.00
; CURRENT APPLICATION NUMBER: US/10/060,857
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 60/264,932
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1050
; TYPE: DNA
; ORGANISM: Coral
US-10-060-857-7

Query Match 99.1%; Score 672.2; DB 14; Length 1050;
Best Local Similarity 99.6%; Pred. No. 1.4e-161;
Matches 674; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	1	ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC	60
Db	289	ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC	348
QY	61	ACCGTGAACGGCCACAGTTCGAGATCGAGGGGAGGGGCGGGCGGCCCTACGAGGC	120
Db	349	ACCGTGAACGGCCACAGTTCGAGATCGAGGGGAGGGGCGGGCGGCCCTACGAGGC	408
QY	121	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTGCCTTGGCGTGGAGATC	180
Db	409	CACAAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTGCCTTGGCGTGGAGATC	468
QY	181	CTGTCCCTCCAGTTCAGTACGGCTTCAAGGTGACGTAAGGACCCCGCCGACATCCCC	240
Db	469	CTGTCCCTCCAGTTCAGTACGGCTTCAAGGTGACGTAAGGACCCCGCCGACATCCCC	528
QY	241	GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGGCGGTGATGAACTTCGAG	300
Db	529	GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGGCGGTGATGAACTTCGAG	588
QY	301	GACGGCGGTGGCGACCGTGAAGGACCTCCCTCCGAGGACCGGTGCTTCTATCTAC	360
Db	589	GACGGCGGTGGCGACCGTGAAGGACCTCCCTCCGAGGACCGGTGCTTCTATCTAC	648
QY	361	AAAGTGAAGTTCTATCGCGTGAATTCCTCCGAGGCGCCCGTGTGAGAGAGAGACC	420
Db	649	AAAGTGAAGTTCTATCGCGTGAATTCCTCCGAGGCGCCCGTGTGAGAGAGAGACC	708
QY	421	ATGGGCTGGAGGCGCTTCCACCGAGCGCTGTACCCCGGAGCGCGTGTGAGAGGGCGAG	480
Db	709	ATGGGCTGGAGGCGCTTCCACCGAGCGCTGTACCCCGGAGCGCGTGTGAGAGGGCGAG	768
QY	481	ATCCAAAGCCCTGAAGTGAAGGACGGCGGCACTACCTGTGAGTTCAAGTCCATC	540
Db	769	ATCCAAAGCCCTGAAGTGAAGGACGGCGGCACTACCTGTGAGTTCAAGTCCATC	828
QY	541	TACATGCCCAAGAGCCCGTGCAGCTGCCCGCTACTACTGTGAGGACACCAAGTGGAC	600
Db	829	TACATGCCCAAGAGCCCGTGCAGCTGCCCGCTACTACTGTGAGGACACCAAGTGGAC	888
QY	601	ATCACCTCCCAACAGGAGCTACACATCTGTGGAGCAGTACGAGCGACCGAGGGCGGC	660
Db	889	ATCACCTCCCAACAGGAGCTACACATCTGTGGAGCAGTACGAGCGACCGAGGGCGGC	948
QY	661	CACCACCTGTTCTCTGTA 677	
Db	949	CACCACCTGTTCTCTGTA 965	

RESULT 5
US-10-006-922-38
; Sequence 38, Application US/10006922


```

; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Pradkov, Arcady F.
; APPLICANT: Lebas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-38

Query Match          98.8%; Score 670.2; DB 14; Length 675;
Best Local Similarity 99.6%; Pred. No. 4.3e-161;
Matches 672; Conservative 0; Mismatches 3; Indels 0; Gaps 0

Qy 1 ATGGCGTCTCTCCGAGAAGCTCATCACGAGTTCATCGCTTCAAGTGGCATGGAGGCG 60
Db 1 ATGGCGTCTCTCCGAGAACGTATCACGAGTTCATCGCGTTCAGGTGGCATGGAGGCG 60

Qy 61 ACGGTGAAGCGCCAGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCGCCCTACGAGGCG 120
Db 61 ACGGTGAAGCGCCAGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCGCCCTACGAGGCG 120

Qy 121 CACAAACCGGTGAAGTTCGAAGGTGACCAAGGGCGGGCGCCCTGCGCTTGGCGTGGACATC 180
Db 121 CACAAACCGGTGAAGTTCGAAGTGAACGACCAAGGGCGGGCGCCCTGCGCTTGGCGTGGACATC 180

Qy 181 CTGTCTCCCGAGTTCAGTAGCGGTCCAGAGTGTACGTGAAGCACACCGCGGCACATCCCG 240
Db 181 CTGTCTCCCGAGTTCAGTAGCGGTTCGAAGGTGTACGTGAAGCACACCGCGGCACATCCCG 240

Qy 241 GACTACAAGAAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATGAATTCGAG 300
Db 241 GACTACAAGAAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATGAATTCGAG 300

Qy 301 GACGGCGGCGTGGCGACCGTGAACCAAGACTTCCTCCCTGAGAGCGGCTGCTTCATCTAC 360
Db 301 GACGGCGGCGTGGCGACCGTGAACCAAGACTTCCTCCCTGAGAGCGGCTGCTTCATCTAC 360

Qy 361 AAGGTGAAGTTCATCGGCGTGAATTCCTCCCTCGAGCGGCCCTGATGAGAGAGAGACC 420
Db 361 AAGGTGAAGTTCATCGGCGTGAATTCCTCCCTCGAGCGGCCCTGATGAGAGAGAGACC 420

Qy 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTTGTACCCCGCGAGCGGCGTGTGAAGGCGCGAG 480
Db 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTTGTACCCCGCGAGCGGCGTGTGAAGGCGCGAG 480

Qy 481 ATCCACAGGCGCCCTGAAGCTGAAGGACGGCGGCCACTACTCTGGTGAAGTTCAGTCCATC 540
Db 481 ATCCACAGGCGCCCTGAAGCTGAAGGACGGCGGCCACTACTCTGGTGAAGTTCAGTCCATC 540

Qy 541 TACATGGCCCAAGAACCGGTGCGAGCTGCCCGGGTACTACTACTACGTGGACACCAAGCTGGAC 600

```

```

Db      541  TACATGCCCAAGAGCCCGTGCAGTGTCCCGGTACTACTACTGTGACGCCAAGCTGGAC 600
Qy      601  ATCACTCTCCCAACACGAGGACTACACCATCGTGGAGCAGTACGAGGCGCACCGAGGGCGCG 660
Db      601  ATCACTCTCCCAACACGAGGACTACACCATCGTGGAGCAGTACGAGGCGCACCGAGGGCGCG 660
Qy      661  CACCACCTGTTCTCTG 675
Db      661  CACCACCTGTTCTCTG 675

RESULT 6
US-10-081-864-13
; Sequence 13, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yarushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-13

```

QY 421 ATGGGTGGAGGCGCTCCACCGAGCGCCTGTACCCCGCGGACGGCGTGTGTAAGGGCGAG 480
Db 421 ATGGGTGGAGGCGCTCCACCGAGCGCCTGTACCCCGCGGACGGCGTGTGTAAGGGCGAG 480
QY 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGCACTACCTGCTGAGTTCAAGTCCATC 540
Db 481 ACCACAAGGCCCTGAAGCTGAAGGACGGCGCACTACCTGCTGAGTTCAAGTCCATC 540
QY 541 TACATGCCCAAGAAGCCCGTGCAGCTGCCCGGTACTACTAGTGGACACCAAGCTGGAC 600
Db 541 TACATGCCCAAGAAGCCCGTGCAGCTGCCCGGTACTACTAGTGGACACCAAGCTGGAC 600
QY 601 ATCACTCCCAACAGAGACTACACCATCTGGAGCAGTACGAGGCGACCCAGGGCGGC 660
Db 601 ATCACTCCCAACAGAGACTACACCATCTGGAGCAGTACGAGGCGACCCAGGGCGGC 660
QY 661 CACCACCTGTTCTCTG 675
Db 661 CACCACCTGTTCTCTG 675

RESULT 7

US-10-315-920-3
; Sequence 3, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Pradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; FILE REFERENCE: FOR THEIR USE
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-3

Query Match 98.6%; Score 668.4; DB 15; Length 678;
Best Local Similarity 99.1%; Pred. No. 1.2e-160;
Matches 672; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 60
QY 61 ACCGTGAAGCGGCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTTACGAGGCG 120
Db 61 ACCGTGAAGCGGCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTTACGAGGCG 120
QY 121 CACAACACCGTGAAGTTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
Db 121 CACAACACCGTGAAGTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
QY 181 CTGTCCCCCAGTTCCAGTACGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 CTGTCCCCCAGTTCCAGTACGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
QY 241 GACTACAAGAGTGTCTTCCCGAGGGGTTCAAGTGGGAGGGCGGTGATGAACTTCGAG 300
Db 241 GACTACAAGAGTGTCTTCCCGAGGGGTTCAAGTGGGAGGGCGGTGATGAACTTCGAG 300
QY 301 GAGGGCGGTGGCGACCGTGAACCCAGGACTCTCTCCCTGACGAGCGGCTTCTATCTAC 360

Db 301 GACGGCGGTGGCGACCGGTGACCCAGGATCTCTCTCCGTGAGGACGGCTGTTCTATCTAC 360
QY 361 AAGTGAAGTTTCATCGCGTGAACCTTCCCTCCGACGGCGCCCGTGTATGCAAGAAAGACC 420
Db 361 AAGTGAAGTTTCATCGCGTGAACCTTCCCTCCGACGGCGCCCGTGTATGCAAGAAAGACC 420
QY 421 ATGGGCTGGAGGCGCTTCACCGAGCGCTTACCCCGCGACGGCGTGTCTGAAAGGGCGAG 480
Db 421 ATGGGCTGGAGGCGCTTCACCGAGCGCTTACCCCGCGACGGCGTGTCTGAAAGGGCGAG 480
QY 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGCACTACCTGCTGAGTTCAAGTCCATC 540
Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGCACTACCTGCTGAGTTCAAGTCCATC 540
QY 541 TACATGCCCAAGAAGCCCGTGCAGCTGCCCGGTACTACTAGTGGACACCAAGCTGGAC 600
Db 541 TACATGCCCAAGAAGCCCGTGCAGCTGCCCGGTACTACTAGTGGACACCAAGCTGGAC 600
QY 601 ATCACTCCCAACAGAGACTACACCATCTGGAGCAGTACGAGGCGACCCAGGGCGGC 660
Db 601 ATCACTCCCAACAGAGACTACACCATCTGGAGCAGTACGAGGCGACCCAGGGCGGC 660
QY 661 CACCACCTGTTCTCTGTA 678
Db 661 CACCACCTGTTCTCTGTA 678

RESULT 8

US-10-081-864-7
; Sequence 7, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; APPLICANT: Lukanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: Methods for Using the Same
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-10-081-864-7

Query Match 98.1%; Score 665.2; DB 15; Length 678;
Best Local Similarity 98.8%; Pred. No. 8e-160;
Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 60
QY 61 ACCGTGAAGCGGCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTTACGAGGCG 120
Db 61 ACCGTGAAGCGGCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTTACGAGGCG 120
QY 121 CACAACACCGTGAAGTTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
Db 121 CACAACACCGTGAAGTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
QY 181 CTGTCCCCCAGTTCCAGTACGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 CTGTCCCCCAGTTCCAGTACGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240

Db 5 TGGCTCTCTCAAGAACGTCATCAAGGAGTTTCATCGCTTCAAGGTGCGCATGGAGGCA 64
Qy 62 CCGTGAACGGCCACAGATTCGAGTCAGAGGGGAGGGGCGCGCCCTACAGAGGCC 121
Db 65 CCGTGAACGGCCACAGATTCGAGTCAGAGGGGAGGGGCGCGCCCTACAGAGGCC 124
Qy 122 ACAACAACCGTGAAGTTGAAGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 181
Db 125 ACAACAACCGTGAAGTTGAAGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 184
Qy 182 TGTGCCCCCAGTTCCAGTACGGCTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 241
Db 185 TGTGCCCCCAGTTCCAGTACGGCTCCAGAGGCTTCCAGAGGCTTCCAGAGGCTTCCAGG 244
Qy 242 ACTACAAGAGCTGCTTCCCTCCCGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 301
Db 245 ACTACAAGAGCTGCTTCCCTCCCGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 304
Qy 302 ACGGCGGCTGCGAGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 361
Db 305 ACGGCGGCTGCGAGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 364
Qy 422 TGGGCTGGAGGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 481
Db 425 TGGGCTGGAGGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 484
Qy 482 TCACAAGGCGCTTGAAGCTGAAGGAGCGGCGGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 541
Db 485 TCACAAGGCGCTTGAAGCTGAAGGAGCGGCGGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 544
Qy 542 ACATGSCCAAGAACGCGCTGAGCTGCGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 601
Db 545 ACATGSCCAAGAACGCGCTGAGCTGCGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 604
Qy 602 TCACCTCCCAACAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 661
Db 605 TCACCTCCCAACAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 664
Qy 662 ACCACCTGTTCTCTGTA 677
Db 665 ACCACCTGTTCTCTGTA 680

RESULT 11
US-10-121-258-23
; Sequence 23, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; FILE REFERENCES: US/10121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DeRed with mammalian
; OTHER INFORMATION: codon usage
US-10-121-258-23

Query Match 97.6%; Score 661.6; DB 15; Length 681;
Best Local Similarity 98.7%; Pred. No. 6.6e-159;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 2 TGGCTCTCTCCGAGAACGTCATCAAGGAGTTTCATCGCTTCAAGGTGCGCATGGAGGCA 61
Db 5 TGGCTCTCTCCGAGAACGTCATCAAGGAGTTTCATCGCTTCAAGGTGCGCATGGAGGCA 64
Qy 62 CCGTGAACGGCCACAGATTCGAGTCAGAGGGGAGGGGCGCGCCCTACAGAGGCC 121
Db 65 CCGTGAACGGCCACAGATTCGAGTCAGAGGGGAGGGGCGCGCCCTACAGAGGCC 124
Qy 122 ACAACAACCGTGAAGTTGAAGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 181
Db 125 ACAACAACCGTGAAGTTGAAGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 184
Qy 182 TGTGCCCCCAGTTCCAGTACGGCTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 241
Db 185 TGTGCCCCCAGTTCCAGTACGGCTCCAGAGGCTTCCAGAGGCTTCCAGAGGCTTCCAGG 244
Qy 242 ACTACAAGAGCTGCTTCCCTCCCGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 301
Db 245 ACTACAAGAGCTGCTTCCCTCCCGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 304
Qy 302 ACGGCGGCTGCGAGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 361
Db 305 ACGGCGGCTGCGAGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 364
Qy 362 AGTGAAGTTTCATCGGCGTGAACCTTCCCTCCAGAGGCGGTGATGAACTTCGAGG 421
Db 365 AGTGAAGTTTCATCGGCGTGAACCTTCCCTCCAGAGGCGGTGATGAACTTCGAGG 424
Qy 422 TGGGCTGGAGGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 481
Db 425 TGGGCTGGAGGCGCTTCCAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 484
Qy 482 TCACAAGGCGCTTGAAGCTGAAGGAGCGGCGGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 541
Db 485 TCACAAGGCGCTTGAAGCTGAAGGAGCGGCGGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 544
Qy 542 ACATGSCCAAGAACGCGCTGAGCTGCGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 601
Db 545 ACATGSCCAAGAACGCGCTGAGCTGCGCGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 604
Qy 602 TCACCTCCCAACAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 661
Db 605 TCACCTCCCAACAGAGGCTTCAAGTGGAGGCGGTGATGAACTTCGAGG 664
Qy 662 ACCACCTGTTCTCTGTA 677
Db 665 ACCACCTGTTCTCTGTA 680

RESULT 12
US-10-161-403-29
; Sequence 29, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30

Publication No. US20020173634A1
GENERAL INFORMATION:
APPLICANT: FRASER JR., MALCOLM J.
APPLICANT: LI, XU
APPLICANT: BEAM, TERESA
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
TITLE OF INVENTION: MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
FILE REFERENCE: 835910-92098
CURRENT APPLICATION NUMBER: US/10/001,189
CURRENT FILING DATE: 2001-10-30
PRIOR APPLICATION NUMBER: 60/244,984
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/244,677
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 45
LENGTH: 6984
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
OTHER INFORMATION: PBAP3-DSRed-orf sequence
US-10-001-189-45

Query Match 97.6%; Score 661.6; DB 14; Length 6984;
Best Local Similarity 98.7%; Pred. No. 7.5e-159;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY	2	TGGCTCTCCGAGAACGTCATCAGGAGTTTCATGGCTTCAAGTGGCGCATGGAGGCA	61
DB	2975	TGGCTCTCCGAGAACGTCATCAGGAGTTTCATGGCTTCAAGTGGCGCATGGAGGCA	3034
QY	62	CGGTGAACCGCCACAGGTTCCGATCGAGGGGAGGGCGGGCGGCGGCTTACGAGGGCC	121
DB	3035	CGGTGAACCGCCACAGGTTCCGATCGAGGGGAGGGCGGGCGGCGGCTTACGAGGGCC	3094
QY	122	ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGGCGGCTTCCCTTGGCTGGGACATCC	181
DB	3095	ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGGCGGCTTCCCTTGGCTGGGACATCC	3154
QY	182	TGTCCTCCCGAGTCCAGTACGGTCCCAAGGTGACGTAAGGACACCGCGGACATCCCGG	241
DB	3155	TGTCCTCCCGAGTCCAGTACGGTCCCAAGGTGACGTAAGGACACCGCGGACATCCCGG	3214
QY	242	ACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAGG	301
DB	3215	ACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAGG	3274
QY	302	ACGGCGGCGTGGCGACCGTGACCGAGTCTCTCCCTGAGGACGCGTCTTCACTACA	361
DB	3275	ACGGCGGCGTGGCGACCGTGACCGAGTCTCTCCCTGAGGACGCGTCTTCACTACA	3334
QY	362	AGGTGAAGTTCATCGGGCGTGAATTCCTCCCGAGCGGCGGCTGATGACGAGAGACCA	421
DB	3335	AGGTGAAGTTCATCGGGCGTGAATTCCTCCCGAGCGGCGGCTGATGACGAGAGACCA	3394
QY	422	TGGGCTGGGAGGCTTCCACGAGGCGCTGACCGGCGGCGGCGGCTGAGAGGGGAGA	481
DB	3395	TGGGCTGGGAGGCTTCCACGAGGCGCTGACCGGCGGCGGCGGCTGAGAGGGGAGA	3454
QY	482	TCCACAGGCGCTTGAAGTGAAGGAGCGGCGGCGGCTTCACTGGTGGAGTTCAAGTCCATCT	541
DB	3455	TCCACAGGCGCTTGAAGTGAAGGAGCGGCGGCGGCTTCACTGGTGGAGTTCAAGTCCATCT	3514
QY	542	ACATGGCCAAAGAGCGGCTGAGTCCCGGCTTCACTACGTTGAGGACACCAAGCTGGACA	601
DB	3515	ACATGGCCAAAGAGCGGCTGAGTCCCGGCTTCACTACGTTGAGGACACCAAGCTGGACA	3574
QY	602	TCACCTCCCAAGAGGAGTACACCATCTGGAGGAGTACGAGCGGCGGCGGCGGCGG	661
DB	3575	TCACCTCCCAAGAGGAGTACACCATCTGGAGGAGTACGAGCGGCGGCGGCGGCGG	3634

QY 662 ACCACCTGTTCTGTGA 677
DB 3635 ACCACCTGTTCTGTGA 3650

RESULT 15
US-10-006-922-35
Sequence 35, Application US/10006922
Publication No. US20020197676A1
GENERAL INFORMATION:
APPLICANT: Lukyanov, Sergey A.
APPLICANT: Fradkov, Arcady F.
APPLICANT: Labas, Yulii A.
APPLICANT: Matz, Mikhail V.
APPLICANT: Tersikh, Alexey
TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
TITLE OF INVENTION: Methods for Using the Same
FILE REFERENCE: CLON-035CIP
CURRENT APPLICATION NUMBER: US/10/006,922
CURRENT FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 09/120,330
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/457,898
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,144
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,477
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/457,556
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/444,338
PRIOR FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 35
LENGTH: 681
TYPE: DNA
ORGANISM: Discosoma species
US-10-006-922-35

Query Match 97.3%; Score 660; DB 14; Length 681;
Best Local Similarity 98.5%; Pred. No. 1.7e-158;
Matches 666; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY	2	TGGCTCTCCGAGAACGTCATCAGGAGTTTCATGGCTTCAAGTGGCGCATGGAGGCA	61
DB	5	TGGCTCTCCGAGAACGTCATCAGGAGTTTCATGGCTTCAAGTGGCGCATGGAGGCA	64
QY	62	CCGTGAACCGCCACGAGTTCCGATCGAGGGGAGGGCGGCGGCTTACGAGGGCC	121
DB	65	CCGTGAACCGCCACGAGTTCCGATCGAGGGGAGGGCGGCGGCTTACGAGGGCC	124
QY	122	ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGGCGGCTTCCCTTGGCTGGGACATCC	181
DB	125	ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGGCGGCTTCCCTTGGCTGGGACATCC	184
QY	182	TGTCCTCCCGAGTCCAGTACGGTCCCAAGGTGACGTAAGGACACCGCGCGACATCCCG	241
DB	185	TGTCCTCCCGAGTCCAGTACGGTCCCAAGGTGACGTAAGGACACCGCGCGACATCCCG	244
QY	242	ACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAGG	301
DB	245	ACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAGG	304
QY	302	ACGGCGGCGTGGCGACCGTGACCGGAGTCTCTCCCTGAGGACGCGTCTTCACTACA	361
DB	305	ACGGCGGCGTGGTGACCGTGACCGGAGTCTCTCCCTGAGGACGCGTCTTCACTACA	364
QY	362	AGGTGAAGTTCATCGGGCGTGAATTCCTCCCTGAGGCGGCGGCGGCGGCGGCGG	421
DB	365	AGGTGAAGTTCATCGGGCGTGAATTCCTCCCTGAGGCGGCGGCGGCGGCGGCGG	424

```
QY 422 TGGGCTGGAGGCGCTCCACCGAGCGCCTGTATACCCCGCGACGGCGTCTGTAAGGGCGAGA 481
Db 425 TGGGCTGGAGGCGCTCCACCGAGCGCCTGTATACCCCGCGACGGCGTCTGTAAGGGCGAGA 484
QY 482 TCCACAGGCGCCTGAAGCTGAAGGACGGCGGCCCACTACCTGGTGGAGTTCAAGTCCATCT 541
Db 485 TCCACAGGCGCCTGAAGCTGAAGGACGGCGGCCCACTACCTGGTGGAGTTCAAGTCCATCT 544
QY 542 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACA 601
Db 545 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACTCCCAAGCTGGACA 604
QY 602 TCACCTCCCAACGAGGACTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 661
Db 605 TCACCTCCCAACGAGGACTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 664
QY 662 ACCACCTGTTCTGTGTA 677
Db 665 ACCACCTGTTCTGTGTA 680
```

Search completed: August 3, 2004, 19:01:41
Job time : 494.957 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 11:02:24 ; Search time 3038.03 Seconds
(without alignments)
6664.376 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggctctctcgaacgt.....gccaacacctgtctctgttaa 678

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_lman:*
23: em_gss_lmus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gssl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.8	12.1	681	12	BI386888
2	79.4	11.7	550	13	BO551001
3	77.8	11.5	629	12	BI384961
4	77.6	11.4	707	12	BI382442

5	77.4	11.4	561	12	BI379061
6	77.4	11.4	577	12	BI386955
7	77.4	11.4	671	12	BI381780
8	77.4	11.4	687	12	BI387917
9	76.4	11.3	459	12	BI386952
10	75.8	11.2	661	12	BI385436
11	75.4	11.1	637	12	BI379468
12	75.4	10.7	591	12	BI382373
13	69.6	10.3	675	12	BI382838
14	69.4	10.2	537	12	BI376909
15	69.4	10.2	826	29	CG448047
16	68	10.0	643	12	BI386699
17	66.4	9.8	670	14	CA258642
18	65.6	9.7	571	12	BI387406
19	65.6	9.7	613	12	BI386976
20	65.2	9.6	629	12	BI388023
21	64.2	9.5	631	13	CA137844
22	64.2	9.5	689	12	BI380055
23	64.2	9.5	702	14	CA258524
24	64.2	9.5	801	29	CG445199
25	64	9.4	568	28	BZ708497
26	64	9.4	629	12	BI380008
27	64	9.4	850	29	CG210258
28	62	9.1	501	13	BU672200
29	61.8	9.1	560	10	BE497059
30	61.8	9.1	725	13	BU672170
31	61.4	9.1	683	14	CA166969
32	61.4	9.1	750	13	BO605927
33	61.4	9.1	799	12	EG310393
34	61.2	9.0	806	29	CC607863
35	61.2	9.0	829	28	BZ643262
36	60.8	9.0	645	13	CA126366
37	60.8	9.0	655	14	CD885812
38	60.6	8.9	624	9	AI948344
39	60.6	8.9	872	14	CB657500
40	60.6	8.9	559	12	BM324075
41	60.4	8.9	607	12	BM323555
42	60.4	8.9	643	12	BM3256281
43	60.4	8.9	650	12	BJ258703
44	60.4	8.9	651	12	BJ257736
45	60.4	8.9	651	12	BJ257736

ALIGNMENTS

RESULT 1
BI386888
LOCUS
DEFINITION
MPMGp531) Branchiostoma floridae cDNA clone MPMGp53119122 5', mRNA sequence.
ACCESSION
BI386888
VERSION
EST.
KEYWORDS
Branchiostoma floridae (Florida lancelet)
SOURCE
Branchiostoma floridae
ORGANISM
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.

BI386888 681 bp mRNA linear EST 26-AUG-2003
BFL26_001591 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGp531) Branchiostoma floridae cDNA clone MPMGp53119122 5', mRNA sequence.

REFERENCE
1 (bases 1 to 681)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, N. and Lehrach, H.
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
Genomes
Genome Res. 13 (6A), 1056-1066 (2003)

JOURNAL
MEDLINE
PUBMED
COMMENT

22683279
12799346
Contact: Panopoulou G
Laboratory 145, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235

Handwritten signature

Db	329	CCTGACCTACGGCGTGCAGTGCTTACGCCGCTACCCCGACCACGTAAGCAGCAGCACTT	388
Qy	246	CAAGAAGCTGTCTTCCCGAGGGCTTCAAGTCGGAGCGCGTGATGAATTCGAGGACGG	305
Db	389	CTTCAAGTCCGCCATGCCCCGAAGGCTACGTCGAGGACGACCATCTTCTTCAAGGACGA	448
Qy	306	CGGCGTGGGACCGTGACCCAGGACTCCTCCCTGCAGACGGCTGTTTCATCTCAAGGT	365
Db	449	CGGCAACTACAAGACCCGCGCGGAGTGAAGTTTCGAGGCGGACACCCTGGTGAACCGCAT	508
Qy	366	GAAGTTTCATCGCGTGACTTCCCTCCGAGCGGCCCGT	404
Db	509	CGAGCTGAAGGGCATCGACTTCAGAGGAGCGGCAACAT	547

RESULT 3	BI384961	LOCUS	BI384961	629 bp	mRNA	linear	EST 26-AUG-2003
DEFINITION	BFLG2 003138 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MEMGP498) Branchiostoma floridae cDNA clone MEMGP498E03107 5', tRNA sequence.						

ACCESSION	FI384961
VERSION	FI384961.1
KEYWORDS	EST.
SOURCE	Branchiostoma floridae (Florida lancelet)
ORGANISM	Branchiostoma floridae
	Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
PREFERENCE	1 (bases 1 to 529)

REFERENCE	1 (bases 1 to 629)
AUTHORS	Panopoulou G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M. and Lehrach, H.
TITLE	New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
JOURNAL	Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE	22683279
PubMed	12793346
COMMENT	Contact: Panopoulou G. Laboratory 145, Dept. Lehrach Max-Planck-Institut fuer Molekulare Genetik Inhnstr. 63-73, D-14195 Berlin, Germany Tel.: +49 30 8413 1235 Fax: +49 30 8413 1128 Email: panopoul@molgen.mpg.de The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: http://www.molgen.mpg.de/amphioxus/ Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (http://www.zmpd.de/)

FEATURES
source

```

convention: BFG or WPMGP498"
/notes=vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
Site_2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSport1 vector using a NotI
(5'-pGACATGTCATGATCGAGCGGCCGCC (T)15-3' and a SalI 5'-
TCGACACGACGATCGG-3' adapters (Gibco BRL)."
```

ORIGIN

Query Match	11.5%;	Score 77.8;	DB 12;	Length 629;
Best Local Similarity	51.7%;	Pred. No. 0.1;	Indels 3;	Gaps 1;
Matches 203;	Conservative	0;	Mismatches 187;	
QY	39	CTTCAAGGTGCGATGGAGGGCACCGGTGAACGGCCACAGGATTCGAGATCGAGGCGGAGGG	98	
Db	104	CCTCGACTTGCACTATATGGCTCCATCAACGGTATGGAGTTCCGACATGTGGCGCGGTGG	163	
QY	99	CGAGGCGCGGCCCTACGAGGGCCACAACCCGTGAAGTTGAAGTGACCAAGGGCGGCGCC	158	
Db	164	GAGTGGCAACCCAAAGACGGATCGCTGAGCGTAAACATGAAGTCTACCAAAGGTG--	220	
QY	159	CTGTGCCCTTCGCGTGGGACATCTGTCCCGCCAGTTCCAGTACGGGTCCCAAGGTGTACGT	218	
Db	221	CTTGCGCGTTTCCTCTGCTGTGTGGCCCGCATCTGGGGTACGGCCACTATCAGTACCT	280	
QY	219	GAAGCACCCGCGCGACATCCCGACTACAAGAAGCTGTCTTCCCGAGGGCTTCAAGTG	278	
Db	281	ACCTTTCCCTGACGGTCCGTGCGCTTTCCAGGCAGCTGTGAACAACGGCGGTATCAAA	340	
QY	279	GGAGCGCGTGATGAATTCGAGAGACGGCGCGTGGCGACCGTGACCCAGAGACTCTCCCT	338	
Db	341	GCATCGCTCTTTCAACTTCGAGACGGTGCCTGCTGACTGCCACCTTCAACACTACTCCTA	400	
QY	339	GCAGGACGGGTGCTTTCATCTACAAGGTGAAGTTTCATCGGGGTGAATTCCTCCCTCCGACGG	398	
Db	401	CAGCGGGCGCAAGATCCATGGAGAGTTTCATCTGTTGGGCAGCTGTTTCCCGCAGCATAG	460	
QY	399	CCCGGTGATCGAAGAAAGACCATGGCTGGGA	431	
Db	461	TCCGGTGATCAACCAACGCGCTGACCGTTTGGGA	493	

RESULT 4
BI382442
LOCUS

BI382442	BI382442	707 bp	mRNA	linear	EST 26-AUG-2003
LOCUS	BI382442	707 bp	mRNA	linear	EST 26-AUG-2003
DEFINITION	BFLG2 000619 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPWGp498) Branchiostoma floridae cDNA clone MPWGp498C08131 5', mRNA sequence.	707 bp	mRNA	linear	EST 26-AUG-2003

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
COMMENT

fax: +49 30 6115-2128
E-mail: panopoul@imgg.mpg.de

The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGGCTTACACTTATGCTCCGGCTCG 3' (M13RSP)

BACKWARD: 5' GCTATTACGACCTGGCGAAGGGGAGTGTG 3' (M13FSP)

Insert Length: 1 Std Error: 0.00

Seq primer: 5'-CCGTCGCGAATCCCGGT-3' pSport3/86

High quality sequence stop: 707.

FEATURES

source

1..707

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMGp498C08131"

/tissue_type="whole embryo"

/dev_stage="5-6 hrs (gastrula stage)"

/lab_host="E.coli, XL1 blue"

/clone_lib="Amphioxus 5-6 hrs cDNA library (Name

convention: BFLG or MPMGP498)"

/note="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5') ;

Site 2: NotI, BamHI, HindIII (3') ; OligodT primed and

directionally cloned in pSport1 vector using a NotI

(5'-pGACTAGTCTAGATCGGAGCGCGCC (T)15-3' and a SalI 5'-

TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.4%; Score 77.6; DB 12; Length 707;

Best Local Similarity 49.3%; Pred. No. 0.11;

Matches 259; Conservative 0; Mismatches 260; Indels 6; Gaps 2;

Qy 39 CTTCAAGTGGCGATGAGGGCCACCGTGAACGCCACGAGTTCGAGATCGAGGGCGAGG 98

Db 133 CTTGACCTGACATCTATGGCTCCATCAACGGTATGGAGTTTGACATGGTGGCGGTG 192

Qy 99 CGAGGGCGCGCTTACGAGGGCCACACACCGTGAAGTTGAAGGTGACCAAGCGCGGCC 158

Db 193 CAGTGGCAACCCAAAGCAAGGTCGCTGACGTAACGTAAGTCTACCAAGGTG- --C 249

Qy 159 CTTGCGCTTCCGCTGGGACATCTCTGTCCTCCCGAGTTCAGTACGGCTCCAGGTGTACGT 218

Db 250 CTTGCGGCTTTCTCTCTGCTGGCGCGCATCTAGGGTACGGCCACTACCAATACCT 309

Qy 219 GAAGCACCCGCGGACATCCCGACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTG 278

Db 310 ACCCTTCCCTGATGGTTCGCTTCCTTCAGGGCATCCGTGAACACGCGGGTATCAAT 369

Qy 279 GAGCGCGTGATGAATTCGAGGACGCGCGGTGGGACCGGTGACCGAGACTTCCTCCCT 338

Db 370 GCATCGCTCTTTCAACTTTGAGGACGCGTGCCTGACTGCCACCTCAACTACTCTCTA 429

Qy 339 GCAGGACGGTGTCTCATCTACAAGGTGAAGTTTCATCGGCGTGAATCTTCCCTCCGAGCG 398

Db 430 CAGCGCGCGGCAAGATCAGGAGAGTTTCTCTGGTGGGACGCTGTTTCCCGACGATAG 489

Qy 399 CCGGTGATCAGAAGAACACATGGCTGGGA- --GGCTCCACCGAGCGCTGTACCC 455

Db 490 TCCGGTGATACCAACACCGCTGACCGCTCTGGACAGAGCGTGGCCAAAGTATGTGCT 549

Qy 456 CCGCGCGGCTGCTGAAGCGCGGAGATCCACAGGCGCTGAAGTGAAGGACGGGCCA 515

Db 550 GTCCGATGACAAGCTTGGCCGAGTTCGTGGACTGGACTTACCTCACCAGAGCGCGGGCG 609

Qy 516 CTACCTGGTGGAGTTCAAGTCCATCTACATGGCCCAAGAGCCCGT 560

Db 610 CTACCGTGCCACGGTGCAGACCAACTTTCACCTTCGCANAGCCCAT 654

RESULT 5

B1379061

LOCUS

DEFINITION

B1379061

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

B1379061 561 bp mRNA linear EST 26-AUG-2003
BFLG1_000692 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498P1084 5',
mRNA sequence.

B1379061

GI:30914228

EST.

Branchiostoma floridae (Florida lancelet)

Branchiostoma floridae

Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

Branchiostoma.

1 (bases 1 to 561)

Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,

Herwig,R., Vingron,M. and Lehrach,H.

New evidence for genome-wide duplications at the origin of

vertebrates using an amphioxus gene set and completed animal

genomes

Genome Res. 13 (6A), 1056-1066 (2003)

22683279

12799346

Contact: Panopoulou G

Laboratory 145, dept.Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr.63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each

cluster is an indicator of the frequency of a transcript in the

analysed library. The cluster size as well as the coordinates of the

rest of the clones assigned to the same fingerprint cluster as the

clone from which the above EST is generated is available at the

amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>

Clones and filters are distributed via the Resource Center/Primary

Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGGCTTACACTTATGCTCCGGCTCG 3' (M13RSP)

BACKWARD: 5' GCTATTACGACCTGGCGAAGGGGAGTGTG 3' (M13FSP)

Insert Length: 1200 Std Error: 0.00

Seq primer: 5'-CCGTCGCGAATCCCGGT-3' pSport3/86

High quality sequence stop: 561.

Location/Qualifiers

1..561

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMGp498P1084"

/tissue_type="whole embryo"

/dev_stage="5-6 hrs (gastrula stage)"

/lab_host="E.coli, XL1 blue"

/clone_lib="Amphioxus 5-6 hrs cDNA library (Name

convention: BFLG or MPMGP498)"

/note="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5') ;

Site 2: NotI, BamHI, HindIII (3') ; OligodT primed and

directionally cloned in pSport1 vector using a NotI

(5'-pGACTAGTCTAGATCGGAGCGCGCC (T)15-3' and a SalI 5'-

TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.4%; Score 77.4; DB 12; Length 561;

Best Local Similarity 52.6%; Pred. No. 0.11;

Matches 193; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

Qy 42 CAAGGTGCGGATGAGGGCCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGGCGGA 101

Db 92 CGACCTTCACATCTTCGGCTCCGTCAACGGCGCGGAGTTCGACTGCTGGAGCGGAAA 151

```

QY 102 GGGCCGCCCCCTACGAGGGGCCAACACCGTGAAGTTGAAGTGACCAAGGGGGGGCCCCCT 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 GGGCAACCCGACGATGGAACCTCGAGACCAAGTGTGAATCCACCCGGGGCG---CCCT 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 162 GCCTTCGCTCGGACATCCTGTCCGCCAGTTCACGACGGCTCCAGGTGTAGTGAA 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 GCCTGCTCCCGCTGCTGATCGGACCCAACTGGGTACGGCTTCTACAGTACTGCC 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 222 GCACCCCGCGCATCCCGCATACAAAGAGTGTCTTCCCGGAGGGCTTCAAGTGGGA 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 CTTCCCTGGCGGCGCTCACCTCCATCCACCGCCATCACGACGAGGGGTACAGGTCA 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 282 GCGCGTGATGAATTCGAGCGGGCGGTGGCCGCGTGGACCGTGGACCGAGGACTCTCCCTGCA 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AGTGTTTCAAGTTTGAAGCGGGAGTGTGAATTCGAATTCGAATTCCTCCGCTACATACGA 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 342 GGAACGCTCTTCATCTACAAGTGAAGTTTCATCGCGCTGAATCTTCCCTCCGAGCGGCC 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 GGGCGGCAAGATCAAGGGGAGTTCACGCTGATAGGGTCAGGTTTCCCTCGCGGGGCC 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 402 CGTGATG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 TGTGATG 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
BI386955 577 bp mRNA linear EST 26-AUG-2003
LOCUS BFL26_001658 Amphioxus 26hr cDNA library (Name convention: BFL26 or
DEFINITION MPMPG531) Branchiostoma floridae cDNA clone MPMPG531M2297 5', mRNA
sequence.
ACCESSION BI386955
VERSION BI386955.1 GI:30921845
SOURCE EST.
ORGANISM Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
1 (bases 1 to 577)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Hewig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, Dept. Leirach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (http://www.rzpd.de)
PCR Primers
FORWARD: 5' CCCAGGCTTACATTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCGAGCTGGGAAGGGGATGTG 3' (M13FSP)
Insert length: 1 Std Error: 0.00
Seq primer: 5'-CCGTCGGAATTCGGGT-3' pSport3/86
High quality sequence stop: 577.

```

```

FEATURES
source
Location/Qualifiers
1..577
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPG531M2297"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention:
BFL26 or MPMPG531)"
/note="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligoat
primed and directionally cloned in pSport1 vector using a
NotI (5'-pCACTAGTTTCAGATCGGACGCGCCGCC (T)15-3' and a
SalI 5'-TCGACCCACGGCTCCG-3' adapters (Gibco BRL)."
ORIGIN
Query Match 11.4%; Score 77.4; DB 12; Length 577;
Best Local Similarity 52.6%; Pred. No. 0.11;
Matches 193; Conservative 0; Mismatches 171; Indels 3; Gaps 1;
QY 42 CAAGTGGCATGGAGGGCACCGTGAACGGCCACGATTCGAGATCGAGGGCGAGCGGA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 CGACCTTCACATCTTCGGCTCCCTCAATGGCGGGAGTTCGACTGGTGGGAGGTGAAA 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 102 GGCGCGCCCTACGAGGGCCACACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCT 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 GGGCAACCCGACGATGGAACCTCGAGACCAAGTGTGAATTCACCCGGGGCG---CCCT 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 162 GCCCTCGCTCGGACATCCTGTCGCCCGAGTTCACAGTACGGCTCCAAGGTGTACGTGAA 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 GCCCTGCTCCCGCTGTGATCGGACCAACCTGGGTACGGCTTCTACCACTACCTGCC 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 222 GCACCCCGCGCATCCCGCATACAAAGAGTGTCTTCCCGGAGGGCTTCAAGTGGGA 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 CTTCCCTCGGGGAGCTTCAACCTTCCAAACCGGCATCACGACGAGGGGTACCAAGTTCA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 282 GCGCGTGATGAATTCGAGGACGGCGGTGGCGACCGTGACCCAGGACTCTCCCTGCA 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 ACGTGTTTCAAGTTTGAAGCGGGAGTGTGAATTCGCTCACTCCGCTACACTACAA 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 342 GGAGCGTGTCTTCACTACAAGGTGAAGTTCATCGGGGTGAATCTTCCCTCCGAGGCC 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 GGGCGGCAAGATCAAGGGGAGTTCAGCTGATCGGTTCAGTTTCCCTCGCGGGCTCC 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 402 CGTGATG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 TGTGATG 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 7
BI381780 671 bp mRNA linear EST 26-AUG-2003
LOCUS BFLG1_003411 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
DEFINITION or MPMPG498) Branchiostoma floridae cDNA clone MPMPG498C047 5',
mRNA sequence.
ACCESSION BI381780
VERSION BI381780.1 GI:30916875
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
1 (bases 1 to 671)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Hewig, R., Vingron, M. and Leirach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
PUBMED

```

COMMENT Contact: Panopoulou G
Laboratory 145, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GTATTATCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGTCGCGAATCCCGGT-3' pSport3/86
High quality sequence stop: 671.
Location/Qualifiers
FEATURES
source
1. .671
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C047"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_hosts="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
/note="vector: pSport1; Site 1: Sall, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTCTAGATCGAGCGGCGGCC (T)15-3' and a Sall 5'-TCGACCCAGCGCTCG-3' adapters (Gibco BRL)."
ORIGIN
Query Match 11.4%; Score 77.4; DB 12; Length 671;
Best Local Similarity 52.6%; Pred. No. 0.12;
Matches 193; Conservative 0; Mismatches 171; Indels 3; Gaps 1;
Qy 42 CAAGTGGCGATGAGGGACCGTGAACCGCCACGAGTTCGAGATCGAGGGCGAGGGCGA 101
Db 71 CGACCTTCACATCTTCGGCTCCGTCATGCGCGGAGTTCGACTGTTGGAGGCGGAAA 130
Qy 102 GGGCGGCGCTTACAGGGGCGACACCGCTGAAGTTGAAGTGAACCAAGGGCGGCGCCCT 161
Db 131 GGGCAACCCGACGATGGAACGCTGGAGACAGTGTGAATCCACCGCGGCGC---CCCT 187
Qy 162 GCGCTTCGCTGGACATCTGTCCCGCCAGTTCACAGTCCAGTGTACGTGAA 221
Db 188 GCGCTTCGCTGGCGCTGTGTGATCGGACCCACCTGGGTGCGGTTTACAGTACCTGCG 247
Qy 222 GCACCGCGCGACATCCCGGACTACAGAAAGTGTCTTCCCGAGGGCTTCAAGTGGGA 281
Db 248 CTTCCCTGGCGGCGCTCACCTTCCTCAACCGCCATCAGCGAGGAGGTACCGATTCA 307
Qy 282 CGCGGTGATGAATTCAGAGACGGCGCGTGGGACCGGTGACCCAGATCTCTCCCTGCA 341
Db 308 ACGTGTGTTCAAGTTTGAAGACCGCGGAGTGTGAATTGCACTTCGCTTACACCTACGA 367
Qy 342 GGACGGCTGCTTACATCTACAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCAGCGGCC 401
Db 368 GGGGCGCAAGATCAAGGGGAGTTCAGCTGATCGGTCAGGTTTCCCTCGCGCGGCGC 427
Qy 402 CGTGATG 408
|||||

Db 428 TGTGATG 434
RESULT 8
B1387917
LOCUS
DEFINITION
B1387917 587 bp mRNA linear EST 26-AUG-2003
BFL26.002620 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531) Branchiostoma floridae cDNA clone MPMGP531N0767 5', mRNA sequence.
B1387917
ACCESSION
B1387917.1 GI:30922753
VERSION
B1387917
KEYWORDS
SOURCE
ORGANISM
Branchiostoma floridae (Florida lancelet)
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma. 687)
1 (bases 1 to 587)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. and Lehraach, H.
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22883279
12799346
Contact: Panopoulou G
Laboratory 145, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GTATTATCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGTCGCGAATCCCGGT-3' pSport3/86
High quality sequence stop: 687.
Location/Qualifiers
FEATURES
source
1. .687
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531N0767"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_hosts="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531)"
/note="vector: pSport1 (Gibco BRL); Site 1: Sall, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTCTAGATCGAGCGGCGGCC (T)15-3' and a Sall 5'-TCGACCCAGCGCTCG-3' adapters (Gibco BRL)."
ORIGIN
Query Match 11.4%; Score 77.4; DB 12; Length 687;
Best Local Similarity 50.7%; Pred. No. 0.12;
Matches 267; Conservative 0; Mismatches 251; Indels 9; Gaps 3;
Qy 42 CAGGTGCGATGAGGGCGACCGGTGAACGCCACGAGTTCGAGATCGAGGGCGAGGGCGA 101

Db 95 CGACATCCACTCCAGGCTCATCAACGGCCACAGTTCGACATGTTGGGCGGAGGAAA 154
 QY 102 GGGCGGCCCCCTACGAGGCGCCACACACCGTGAAGTTGAAGTGACCAAGGCGGCCCT 161
 Db 155 AGGCGACCCGAAACGCGGCTCGCTGGTACACACAGGAAATCCACCAAGGCT---CCCT 211
 QY 162 GCCCTTCGCTGGGACATCCTCTCCCGCAGTTCCAGTACGCTCCCAAGGTGTACGTGAA 221
 Db 212 GAAGTTCTCTCCCTACTTATGATATCCCGCAGTCCCGTACGCTGTTACCTACCTCCC 271
 QY 222 GCACCCCGCGCATCCCGACTACAAG---AAGTGTCTCTTCCCGGAGGCTTCAAGTG 278
 Db 272 CTACCGGAGGACCGCTCGCTTCCAGGCGACCATGTTGGAAGAACAGAGGTATGCAGT 331
 QY 279 GAGGCGCTGATGAATTCGAGAGCGGCGCTGCGACCGTGCAGTCCCGAGGACTCTCCCT 338
 Db 332 CCATCGCGTTCGACTTTGAAGAGCGGAGGCAAGTGTCCATCGAGTTTAAGTACTCTTA 391
 QY 339 GAGGACGCGCTGCTTCATCTACAAGGTGAAGTTTCATCGGCGTGAATCTTCCCTCCGACGG 398
 Db 392 CGAGGGTTCCTATCAAGGCGGACATGAAGTTTACGGGACCGGTTTCCCTGAGGACGG 451
 QY 399 CCGGTGATGAGAGAACACATCGGCTGGGAGGCGCTCC---ACGAGCGCTGTATCC 455
 Db 452 GCGGTTCATGACGACGACGATGTCGACGAGGCGCTGCGTGTCCAAAGAACACCTACCT 511
 QY 456 CGCGACGCGCTGCTGAAGGCGGAGATCCACAGCGCTGAGCTGAAGGACGCGGCGCA 515
 Db 512 TACGACACACATCGTGGACAGTTCGCTGGACTTACACCTGCGAGATGGAGCG 571
 QY 516 CTACCTGTGTGAGTTCATCATCTCATGTCATGATGTCGCAAGGCGCGCTGC 562
 Db 572 CTCAAGGCGCAAGTGACGAGCGCACTACATCTTCGCGAAGCGCTTTC 618

RESULT 9

BI386952 459 bp mRNA linear EST 26-AUG-2003
 LOCUS BFL26_001655 Amphioxus 26hr cDNA library (Name convention: BFL26 or
 MPMGP531) Branchiostoma floridae cDNA clone MPMGP531M0997 5', mRNA
 sequence.

ACCESSION

BI386952.1 GI:30921843
 EST.
 Branchiostoma floridae (Florida lancelet)
 Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

REFERENCE

1 (bases 1 to 459)
 Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
 Herwig,R., Vingron,M. and Lehrach,H.
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes

Genome Res. 13 (6A), 1056-1066 (2003)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Panopoulou G
 Laboratory 145, dept.Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihnestr.63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the

clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rspd.de>)
 PCR Primers
 FORWARD: 5' CCCAGACTTTACACTTTATGTTCCGGCTCG 3' (M13RSP)
 BACKWARD: 5' GGTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
 Insert Length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGCGAATTCGCGGT-3' pSport3/86
 High quality sequence stop: 459.

FEATURES

source
 1. 459
 Location/Qualifiers
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMGP531M0997"
 /tissue_type="whole embryo"
 /dev_stage="26 hrs (neutula stage)"
 /lab_host="E.coli, XL1 blue"
 /clone_lib="Amphioxus 26hr cDNA library (Name convention:
 BFL26 or MPMGP531)"
 /notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
 EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligo dT,
 primed and directionally cloned in pSport1 vector using a
 NotI (5'-pGACTAGTTCATGATCGGAGCGGCCCC (T)15-3' and a
 SalI 5'-TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.3%; Score 76.4; DB 12; Length 459;
 Best Local Similarity 52.3%; Pred. No. 0.15;
 Matches 192; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 42 CAAGGTCCGATGAGGCGCACCGTGAACCGCCACGAGTTCGAGATCGAGGCGGAGGCGGA 101

Db 79 CGACCTTCACATCTTCGGCTCCGTCATGCGCGGAGTTCGACCTGTGTGGAGCGGAAA 138

QY 102 GGGCGGCGCTAGAGGGCCACACACCGTGAAGTGAAGTGAAGGCGGCGGCCCT 161

Db 139 GGGCAACCCGACGATGGAACCGTGGAGACAGTGTGAATCCACCGGGGCG---CCCT 195

QY 162 GCCCTTCGCTGGGACATCCTGTCCCGCCAGTTCACGATACGGCTCCAAAGGTGTACGTGAA 221

Db 196 GCCCTGCTCCCGCTGCTGATCGAGCCCAACCTGGGTACGGCTTCTACCACTACCTGCC 255

QY 222 GCACCCGCGGACATCCCGGATACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGGA 281

Db 256 CTTCCCTGGGGCGGCTCGCCCTTCCAAACAGCCATCACGAGCGAGGTACCAAGTTCA 315

QY 282 GCGGTGATGAATTCGAGGACGCGCGGTGGCGACCGTGCACCCAGGACTCCTCCCTGCA 341

Db 316 ACGTGTTCAGTTTGAAGACCGGGAGTGTGAATTGCACTTCGCTACACTACNA 375

QY 342 GGACGGTGTTCATCTPACAAGGTGAAGTTTCATCGGCGTGAATTCCTCCCTCCGACGCC 401

Db 376 GGGCGGTAAAGATCAAGGGGAGTTCAGCTGATCGGTTCCTCCCGCGCGGCC 435

QY 402 CGTGATG 408

Db 436 TGTGATG 442

RESULT 10

BI385436

LOCUS

DEFINITION

MPMGP531

sequence.

ACCESSION

BI385436

VERSION

BI385436.1

GI:30920419

KEYWORDS

SOURCE

ORGANISM

Branchiostoma floridae (Florida lancelet)

Branchiostoma floridae

Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

BI385436 561 bp mRNA linear EST 26-AUG-2003
 BFL26_000139 Amphioxus 26hr cDNA library (Name convention: BFL26 or
 MPMGP531) Branchiostoma floridae cDNA clone MPMGP531H0446 5', mRNA
 sequence.

Branchiostoma.
1 (bases 1 to 661)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GGTATACCGACCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGATTCCTCGGT-3' pSport3/86
High quality sequence stop: 661.
Location/Qualifiers
1. 661
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531H0446"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neural stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention:
BFL26 or MPMGP531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT,
primed and directionally cloned in pSport1 vector using a
NotI (5')-pGATGTTTATGATCGGAGCGGCCCC (T)15-3' and a
SalI 5'-TCGACCCACGGTCCG-3' adapters (Gibco BRL)."
ORIGIN
Query Match 11.2%; Score 75.8; DB 12; Length 661;
Best Local Similarity 50.5%; Pred. No. 0.2;
Matches 266; Conservative 0; Mismatches 252; Indels 9; Gaps 3;
QY 42 CAAGGTGCGATGAGGCGACCGTGAACCGCCACGAGTTCGAGATCGAGGGCGAGGCGA 101
Db 61 CGACATCCACCTCCAGCGCTCCATCAACGGCCACGAGTTCGACTGTTGGGGGAGGACA 120
QY 102 GGCGCCCGCTAGAGGGCCACAAACCCGTGAGGTGACCAAGGGCGGCCCT 161
Db 121 AGCGACCGAAGCGCGCTCGGTGTGACCAAGCGAATCCACCGAGGT---CCCT 177
QY 162 GCCTTTCCGCTGGGACATCTGTCCCGCCAGTTTCCAGTACGGCTCCAAAGGTGTACGTGAA 221
Db 178 GAAGTTCTCTCCCTACTTTGATGATCCCCACCTCGGGTACGGGTACTACCAAGTACCTCCC 237
QY 222 GCACCCCGCGACATCCCGCATCAAG---AGCTGTCTTCCCCGAGGGCTTCAAGTG 278
Db 238 CTACCCGGAGCGACCTCGCTTTCCAGGCCACCATGTTGGGAAGGATCAGGTGTGACGT 297

QY 279 GGAGCGGTGATGAATTCGAGGACGGCGGTGGCGACCGTGCACCGAGCTCCTCCCT 338
Db 298 CCATCGGGTGTTCGACTTTGAAGACGGGACGAGCTGTCCATCGAGTTTAAATCTCTTA 357
QY 339 GCAGGACGGGTGCTTCATCTACAAAGGTGAAGTTTCATCGCGGTGAACATTCCTCCCGACGG 398
Db 358 CGAGGTTCCATATCAAGCGCGACATGAAGTTTCAGCGGAACCGGTTTCCCTGAGGACGG 417
QY 399 CCCGCTGATGCAGAAGAACCATGGCTGGAGGCTCC---ACCGAGCGCTGTACCC 455
Db 418 GCCGCTCATGACCGCCAGATTGTGCAACCGAGCGGTGCTGCTTCCAGAACACCTACCT 477
QY 456 CCGCGACGGCTGCTGAAGGGCGAGATCCCAAGGCCCTGAAGCTGAAGGACGCGCGCCA 515
Db 478 TAACGACACACCATCTGTGACACGCTTCGACTGACTTACACTGCAGATGGAGCG 537
QY 516 CTACTGCTGGAGTTCAGTCCATCTATCATGGCCGCAAGACCGCGTGC 562
Db 538 CTACAAGGCCCAAGTACGAGCCACTACATCTTCGGCAAGCCCTTTC 584

RESULT 11

LOCUS B1379468
DEFINITION B1379468 637 bp mRNA linear EST 26-AUG-2003
ORIGIN BFLG1.001099 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMGP498) Branchiostoma floridae cDNA clone MPMGP49801918 5',
mRNA sequence.
ACCESSION B1379468
VERSION B1379468.1 GI:30914617
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
REFERENCE 1 (bases 1 to 637)
AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Herwig, R., Vingron, M. and Lehrach, H.
TITLE New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE 22683279
PUBMED 12799346
COMMENT Contact: Panopoulou G
laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GGTATACCGACCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1200 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGATTCCTCGGT-3' pSport3/86
High quality sequence stop: 637.
Location/Qualifiers
1. 637
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"

JOURNAL
MEDLINE
PUBMED
COMMENT

Genome Res. 13 (6A), 1056-1066 (2003)
22683279
1279346
Contact: Panopoulou G
Laboratory 145, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Amphioxus and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACATTTATCTCCGGCTCG 3' (ML3RSP)
BACKWARD: 5' GCTATTACCCAGCTGGGAAAGGGGATGTG 3' (ML3FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGTCGGAATCCCGGGT-3' p8port3/86
High quality sequence stop: 591.

Location/Qualifiers
1..591
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C1177"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name
convention: BFLG or MPMGP498)"
/note="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSport1 vector using a NotI
(5'-PGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-
TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

FEATURES
source

Query Match 10.7%; Score 72.6; DB 12; Length 591;
Best Local Similarity 51.8%; Pred. No. 0.55;
Matches 190; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

QY 42 CAAGGTGGGATGAGGCGACCCGTGAACGCGCCAGCTCGAGTCGAGGCGGAGGCGGA 101
Db 91 CGACCTTTCACATCTTCGGCTCCGTCATGCGCGGAGTTCGACCTGGTGGGAGCGGAAA 150
QY 102 GGGCGGCCCTTACGAGGCGCACACACCGTGAAGTTGAAGTTCACCAAGGCGGCCCTT 161
Db 151 GGGCAACCGGATGATGAGACCGTGGAGACCATGTGAAATCCACCGGGGCG--CGCT 207
QY 162 GCGCTTGGCTGGGACATCTCTGTCCTCCCGAGTTCAGTCCAGTCCAGGTCGAGTGTACGTGAA 221
Db 208 GCGCTTCTCCCGCTGTGATCGGACCCCAACCTGGGTACGGCTTCTACCAAGTACCTGCC 267
QY 222 GCACCGCGCGACATCCCGACTACAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGGA 281
Db 268 CTTCCCTGGCGCGCTTACCCCTTCCAAACCGCCATCAGGACGAGGGTACCAAGTTCA 327
QY 282 GCGGTGATGAATTCGAGGACGCGCGGTGGGACCGGTGACCCAGGACTCTCCCTGCA 341
Db 328 CCGTGTGTCAAGTTTGAAGACGAGGATTCTGAATTGCAATTCGCGTACACTACGA 387
QY 342 GAGCGGTGCTTCATCTACAGGTGAATTCATCGGCGTGAACCTTCCCTCCGACGCC 401
Db 388 GGGCGGCAAGATCAAAAGGGGAGTTCAGCTGATCGGTCAGGTTTCTCTCCGCGCGGCC 447

/clone="MPMGp498C1177"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name
convention: BFLG or MPMGP498)"
/note="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSport1 vector using a NotI
(5'-PGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-
TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.1%; Score 75.4; DB 12; Length 637;
Best Local Similarity 50.5%; Pred. No. 0.22;
Matches 265; Conservative 0; Mismatches 251; Indels 9; Gaps 3;

QY 42 CAAGGTGGGATGAGGCGACCCGTGAACGCGCCAGCTCGAGTTCGAGTTCGAGGCGGAGGCGGA 101
Db 78 CGAGATCCACCTCCAGGTTCCATCAACGCGCCAGGTTGCGACCTGGCTGGTGGGAAA 137
QY 102 GGGCGGCCCTTACGAGGCGCACACACCGTGAAGTTGAAGTTCACCAAGGCGGCCCTT 161
Db 138 AGGCGACCCAAACGCGCGTCTGTCGACCAACGCAATCCACCCAG--GGTCCCT 194
QY 162 GCGCTTGGCTGGGACATCTCTCCCGAGTTCAGTACGCTCCAAAGTGTAGTGAA 221
Db 195 GAAGTCTCTCCGCTTGTGATCCCGACCTCGGTACGGTACTACCACTACCTCC 254
QY 222 GCACCGCGCGACATCCCGACTACAAG--AAGTCTCTTCCCGAGGCGGTTCAGGTG 278
Db 255 CTACCGGAGCGACCTCGCTTCCAGGCGCACCATGTTGGAAGGATCGGGTATACAGT 314
QY 279 GAGCGCGGTGATGACTTCGAGGCGCGCGTGGCGCGGATGACCCAGGACTCTCTCC 338
Db 315 CCACCGCGTGTGATCTCGAAGCGGAGGAGCTGTCATCGAGTTCAGTACTCTCTA 374
QY 339 GAGGAGCGGCTGTTTCATCTACAAGGTGAAGTTTCATCGGCGTGAATTCCTCCCGACGG 398
Db 375 CGAGGCTTCCATATTAAGCGCGACATGAAGTTTCAGGGAACCGGTTTCCCTGAGGATGG 434
QY 399 CCGCGTGTGAGAGAGAACATCGGCTGGAGGCGCTCC--ACGAGCGCTTACGCC 455
Db 435 GCGGCTCATGACGAGCGAGTGTTCGACGAGGCGCTGCTGTCCAAAGAACCTTACT 494
QY 456 CCGCGAGCGCGTCTGAAGGCGGAGATCCAAAGGCGCTGAAGCTGAAGGAGCGGCGCA 515
Db 495 CACGCAACACCATCTGTCGACAACTTCGACTGACTTACACCTGCGAATGGGAGCG 554
QY 516 CTACCTGTGTGAGTTCAAGTTCATCTACATGCGCCAAAGAGCGCGT 560
Db 555 CTACAGGCGCGAGTCAAGGCGCACTACATCTTTCGACAAAGCGCTT 599

RESULT 12
BI382373
LOCUS
DEFINITION
BFLG2 000550 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498C1177 5',
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BI382373
BI382373.1 GI:30917444
EST.
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
1 (bases 1 to 591)
Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
Herwig,R., Vingron,M. and Lehrach,H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes

QY 402 CTTGATG 408
 Db 448 TGTGATG 454

RESULT 13
 BI382638 675 bp mRNA linear EST 26-AUG-2003
 LOCUS BFLG2_000815 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
 or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498F0258 5',
 mRNA sequence.
 ACCESSION BI382638
 VERSION BI382638.1 GI:30917704
 KEYWORDS EST.
 SOURCE Branchiostoma floridae (Florida lancelet)
 ORGANISM Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 Branchiostoma.
 1 (bases 1 to 675)
 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
 Herwig, R., Vingron, M. and Lehrach, H.
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes
 Genome Res. 13 (6A), 1056-1066 (2003)
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Innestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoul@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the
 clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rzpd.de>)
 PCR Primers
 FORWARD: 5' CCCAGCTTACATTTATGCTCCGCTCG 3' (M13RSP)
 BACKWARD: 5' GTATTACCCAGCTGGGAAAGGGGATGTG 3' (M13FSP)
 Insert length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGGAAATCCCGGGT-3' pSport3/86
 High quality sequence stop: 675.
 Location/Qualifiers
 1..675
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMPGp498F0258"
 /tissue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="amphioxus 5-6 hrs cDNA library (Name
 convention: BFLG or MPMPGp498)"
 /notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
 Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
 directionally cloned in pSport1 vector using a NotI
 (5'-pGACTAGTCTAGATCGGCGCGGCC (T)15-3' and a SalI 5'-
 TCGACCCACGCGCTCG-3' adapters (Gibco BRL)."

Query Match 10.3%; Score 69.6; DB 12; Length 675;
 Best Local Similarity 49.7%; Pred. No. 1.5;

ORIGIN

Matches 261; Conservative 0; Mismatches 255; Indels 9; Gaps 3;
 QY 42 CAAGGTGCGCATGGAGGCCACCGTGAACGGCCACAGAGTTGAGATCGAGGGCGAGGCGA 101
 Db 106 CGACATCCACCTTCACGGCTCCATCAACGGCCACAGATTCACATGCTGGGGGAGGAAA 165
 QY 102 GGGCGCCCTTACGAGGCCACACCGTGAAGTTGAGGTGACCAAGGGCGGCCCT 161
 Db 166 AGGCGACCCGACCGCGCTCGCTGGTGACCCACAGGAAATCCACCAAGGT---CCCT 222
 QY 162 GCCCTTCCCTTGGGACATCTTGTCCCGCCAGTTCCAGTACGGCTCCCAAGGTGTACGTGAA 221
 Db 223 GAAGTTCTCCCTTACTTGATGATCCCGCCACCTCGGTAGGGTACTACAGTACCTCCC 282
 QY 222 GCACCCCGCCACATCCCGGACTACAAGAGCTGCTCTTCCCCGA---GGCTTCAAGTG 278
 Db 283 CTACCCGGACGACCTCGCTTCCAGGTCTCCATGTGGAAGGATCGGGGTATGAGT 342
 QY 279 GGAGCGGTGATGAATTCGAGGACGGCGGTGGCGACCGTGACCCAGGACTCTCTCCCT 338
 Db 343 CTACCGGTGTTCCGACTTTGAGACGGGAGCAAGCTGACTACCGAGTTTAAGTACTCTA 402
 QY 339 GCAGACGGCTGCTTCTATCTACAAGGTGAAGTTTCATCGGGGTGAATTTCCCTCCCGACGG 398
 Db 403 CGAGGGTTCCCATATCAAGGCCGACATGAAGCTGATGGGAAGCGGTTTCCCTGACGACGG 462
 QY 399 CCCGTGATGCAGAGAAGACCATGGGCTGGAGGCTCC---ACCGAGCGCTGTATCCC 455
 Db 463 CCCAGTATGACCGACGAGATTGTGACCGAGCGGTGCTGCTCCAAAGACGATATCT 522
 QY 456 CCGGACGGGTGCTGTAAGGCGGAGATCCACAAGGCCCTGAAGCTGAAGACGCGGCCA 515
 Db 523 TACAACAACACCATCGTGACAGCTTCGACTGGAGTTACAACTGACAGATGGGAAGCG 582
 QY 516 CTACCTGGTGGAGTTCAAGTCCATCTACATGCGCCAGAGCCCTT 560
 Db 583 CTACAAGCGCGAGTGTCTNAGCCACTATCTTCGACAAGCCCTT 627

RESULT 14
 LOCUS BI376909 537 bp mRNA linear EST 26-AUG-2003
 DEFINITION BFLG3_000712 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
 or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498C2412 5',
 mRNA sequence.
 ACCESSION BI376909
 VERSION BI376909.1 GI:30911858
 KEYWORDS EST.
 SOURCE Branchiostoma floridae (Florida lancelet)
 ORGANISM Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 Branchiostoma.
 1 (bases 1 to 537)
 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
 Herwig, R., Vingron, M. and Lehrach, H.
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes
 Genome Res. 13 (6A), 1056-1066 (2003)
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Innestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoul@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the
 clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rzpd.de>)
 PCR Primers
 FORWARD: 5' CCCAGCTTACATTTATGCTCCGCTCG 3' (M13RSP)
 BACKWARD: 5' GTATTACCCAGCTGGGAAAGGGGATGTG 3' (M13FSP)
 Insert length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGGAAATCCCGGGT-3' pSport3/86
 High quality sequence stop: 675.
 Location/Qualifiers
 1..675
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMPGp498F0258"
 /tissue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="amphioxus 5-6 hrs cDNA library (Name
 convention: BFLG or MPMPGp498)"
 /notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
 Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
 directionally cloned in pSport1 vector using a NotI
 (5'-pGACTAGTCTAGATCGGCGGCC (T)15-3' and a SalI 5'-
 TCGACCCACGCGCTCG-3' adapters (Gibco BRL)."

Query Match 10.3%; Score 69.6; DB 12; Length 675;
 Best Local Similarity 49.7%; Pred. No. 1.5;

ORIGIN

cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers
FORWARD: 5' CCCCAGGTTTACACTTATGCTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTAGCCAGCTGGCAAGAGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 537.

Location/Qualifiers
1. 537
/organism="Branchiostoma floridae"
/mol_type="rRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C24112"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGp498)"
/note="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligodt primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGGCTCGG-3' adapters (Gibco BRL)."

ORIGIN		Query Match	10.2%; Score 69.4; DB 12; Length 537;
		Best Local Similarity	51.2%; Pred. No. 1.6;
		Matches	188; Conservative 0; Mismatches 176; Indels 3; Gaps 1;
QY	42	CAAGGTGCGCATGGAGGCGACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGA	101
Db	74	CGACCTTCAATCTTCGGCTCCGTCAATGTGCGGAGTTCGACCTGTTGGAGAGGAGAAC	133
QY	102	GGGCGCCCGCTTACGAGGCGCAACAACCGTGAAGTTGAAGGTGACCAAGGGCGGCCCTT	161
Db	134	GGGCAACCCGACCATGGAACCGCTGGAGACCAAGTGTGAATCCACCCGGGGGGG---CCCT	190
QY	162	GCCTTCGCTGGGATCCTGTCCCGCCAGTTCAGTACGGCTCCCAAGGTGTAGTGAA	221
Db	191	GTCTGTCTCCCGCTGCTGATCGGACCCAACTTGGGTACGGCTTACAGTACTGCTGC	250
QY	222	GCACCCCGCGCATCCCGACTACAAGAGTGTCTCTTCCCGAGGGCTTCAAGTGGGA	281
Db	251	CTTCCCTGGCGGCGCTCCGCTTCCAAACCGCCATATTGACGGAGGTACAGATTCA	310
QY	282	GCGCGTGATGAATTCGAGACGGCGGCTGGCGACCGGTGACCCAGGACTCTCTCCCTGA	341
Db	311	CCGTGTGTTCAGTTTGAAGACGGAGGAATGCTGAAATTCGAATTCCTGATACCTACGA	370
QY	342	GGACGGGTCTTCATCTACAAGTGAAGTTTCATCGGGGTGAATTCCTCCCGACGGCCC	401
Db	371	GGCGGCGAAGATCAAGGGGAGTTCACCTGTCTGGGTACAGTTTCCTTCGCGGGGGCC	430
QY	402	CGTGATG 408	
Db	431	TGTGATG 437	

RESULT 15
CG448047
LOCUS
DEFINITION
OGVGS13TH ZM 0.7 1.5 KB Zea mays genomic clone ZM8Ma0478D01,
genomic survey sequence.
ACCESSION
CG448047
VERSION
GSS.
KEYWORDS
Zea mays

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 826)
Whitlaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGVGS13TV
Contact: Cathy Whitelaw

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org

Seq primer: TR
Class: sheared ends.

Location/Qualifiers
1. 826
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM8Ma0478D01"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pESK-; Site 1: HindIII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN		Query Match	10.2%; Score 69.4; DB 29; Length 826;
		Best Local Similarity	50.4%; Pred. No. 1.7;
		Matches	203; Conservative 0; Mismatches 191; Indels 9; Gaps 1;
QY	259	TTCCCGGAGGCTTCAAGTGGAGCGCGTGTATGAACATTCGAGACGCGGGTGGCGACC	318
Db	170	TTCCGCAAGGTGTTCAGAGGGGTCTGTCCAAAGTCCCGCACGGAGGTCCCGTGAAGCGG	229
QY	319	GTGACCCAGGACTCCTCCCTGCAGAGCGGTCTTCACTACAGGTGAAGTTTCATCGGC	378
Db	230	GTGTCACAGACTCCACGAGGGCTCAGGAGTTTCACTCCGAGGTCTGTCAGATCGGC	289
QY	379	GTGAACCTTCCCTCCGAGCGCCCGTGTATGCAGAAAGACCATGCGCTGGAGGCGCTCC	438
Db	290	CACCTCCGCGCACCGCAACCTGTCAGCTGTGGGTACTGCCGAGAGAGG-----	342
QY	439	ACCGAGCGCTGTACCCCGGAGCGGTCTGAGGGCGAGATCCACAGGCCCTGAAG	498
Db	343	--CGAGCTGTCTGTGTCTACGACTACATGCCCAACGGCAGCTCGACAGCACTGCAC	400
QY	499	CTGAAGGACGGCGCCACTTACCTGTGGAGTTCAAGTCCATCTACATGGCCCAAGAGCCC	558
Db	401	GGCGAGGACGCAAGCCGCTCTGGAGTGGCGGCGAGGTTCCAGATCGTCAAGACGTG	460
QY	559	GTGCAGTGTCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACTCCCAACAGAG	618
Db	461	GGTCCGGCTCTTCTTACTCTCCAGAGAAAGTGGAGGAGGTGTGTCGTCCACCGGACGTC	520
QY	619	GACTACACCATCTGGAGCAGTACGAGCGCACCGAGGGCGGCC 661	
Db	521	AAGGCCAGCAACGTCTCTCGACGGCGCATGTGTGCCACC 563	

Search completed: August 3, 2004, 15:47:59
Job time : 3043.03 secs

Blank Sheet U.S. 'PTD

Ab01622 Discosoma
Aag6509 Anthozoa
Ab08834 Yeast opt
Aae2826 Discosoma
Aae2883 Discosoma
Aae1754 Discosoma
Aae1827 Discosoma
Aae3496 Discosoma
Aae24126 Discosoma
Aae00918 Discosoma
Ab08821 Autofluor
Ab08822 Autofluor
Aae28924 Discosoma
Ab00929 Discosoma
Abw00938 Discosoma
Abw00930 Discosoma
Aag6510 Anthozoa
Ab08835 Yeast opt
Aae24109 Discosoma
Aae28922 Discosoma
Abw00936 Discosoma
Abw00931 Discosoma
Abw00932 Discosoma
Abw00935 Discosoma
Abw00939 Discosoma
Abw00941 Discosoma
Abw00940 Discosoma
Abw00933 Discosoma
Abw00934 Discosoma
Aap56678 Mammalian
Aae28923 Discosoma
Aap56685 Discosoma
Aap70039 Colour Fa
Aap56684 Discosoma

11 1191 96.3 225 3 AAB01622
12 1191 96.3 225 4 AAG6509
13 1191 96.3 225 5 ABB08834
14 1191 96.3 225 5 AAE2826
15 1191 96.3 225 5 AAE2883
16 1191 96.3 225 5 AAE1754
17 1191 96.3 225 5 AAO1827
18 1191 96.3 225 6 AAE3496
19 1191 96.3 225 7 ADC24126
20 1191 96.3 225 7 ABW00918
21 1191 96.3 487 5 ABB08821
22 1191 96.3 506 5 ABB08822
23 1191 96.3 547 5 ABB08823
24 1189 96.0 225 5 AAE28924
25 1188 96.0 225 7 ABW00929
26 1188 96.0 225 7 ABW00938
27 1187 96.0 226 4 AAG6510
28 1187 96.0 226 5 ABB08835
29 1187 96.0 242 7 ADE24109
30 1187 96.0 225 5 AAE28922
31 1186 95.9 225 7 ABW00936
32 1186 95.9 225 7 ABW00931
33 1186 95.8 225 7 ABW00932
34 1185 95.8 225 7 ABW00935
35 1185 95.8 225 7 ABW00939
36 1185 95.8 225 7 ABW00941
37 1184 95.7 225 7 ABW00940
38 1184 95.7 225 7 ABW00933
39 1183 95.6 225 7 ABW00934
40 1183 95.6 225 6 AAP56678
41 1182 95.6 225 5 AAE28923
42 1181 95.5 240 6 AAP56685
43 1179 95.3 225 5 AAP70039
44 1177 95.1 225 5 AAP56684
45 1176 95.1 240 6 AAB56684

11 1191 96.3 225 3 AAB01622
12 1191 96.3 225 4 AAG6509
13 1191 96.3 225 5 ABB08834
14 1191 96.3 225 5 AAE2826
15 1191 96.3 225 5 AAE2883
16 1191 96.3 225 5 AAE1754
17 1191 96.3 225 5 AAO1827
18 1191 96.3 225 6 AAE3496
19 1191 96.3 225 7 ADC24126
20 1191 96.3 225 7 ABW00918
21 1191 96.3 487 5 ABB08821
22 1191 96.3 506 5 ABB08822
23 1191 96.3 547 5 ABB08823
24 1189 96.0 225 5 AAE28924
25 1188 96.0 225 7 ABW00929
26 1188 96.0 225 7 ABW00938
27 1187 96.0 226 4 AAG6510
28 1187 96.0 226 5 ABB08835
29 1187 96.0 242 7 ADE24109
30 1187 96.0 225 5 AAE28922
31 1186 95.9 225 7 ABW00936
32 1186 95.9 225 7 ABW00931
33 1186 95.8 225 7 ABW00932
34 1185 95.8 225 7 ABW00935
35 1185 95.8 225 7 ABW00939
36 1185 95.8 225 7 ABW00941
37 1184 95.7 225 7 ABW00940
38 1184 95.7 225 7 ABW00933
39 1183 95.6 225 7 ABW00934
40 1183 95.6 225 6 AAP56678
41 1182 95.6 225 5 AAE28923
42 1181 95.5 240 6 AAP56685
43 1179 95.3 225 5 AAP70039
44 1177 95.1 225 5 AAP56684
45 1176 95.1 240 6 AAB56684

ALIGNMENTS

RESULT 1
AAE28837
ID AAE28837 standard; protein; 225 AA.
XX AC AAE28837;

XX XX 27-DEC-2002 (first entry)
XX DE Discosoma sp. drFP583 (NFP-6) mutant protein, E5-NA.

XX KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; drFP583 protein; NFP-6; muten;
mutant.
XX OS Discosoma sp.
XX OS Synthetic.

XX FH Key Location/Qualifiers
XX FT Misc-difference 2 /note= "Wild type Arg substituted with Ala"
XX FT Misc-difference 5 /note= "Wild type Lys substituted with Glu"
XX FT Misc-difference 9 /note= "Wild type Lys substituted with Thr"
XX FT Misc-difference 105 /note= "Wild type Val substituted with Ala"
XX FT Misc-difference 197 /note= "Wild type Ser substituted with Thr"
XX MO200268459-A2.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:23:50 ; Search time 48.2188 Seconds
(without alignments)
7945.752 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggcctctctcgagacgt.....gccaccaccttctctgttaa 678

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_29072004_150505_25453/app_query.fasta_1.1742
-DB=A_Geneseq_29Jan04 -OPMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 -CGN_1_1_128/runat_29072004_150505_25453 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSP=LOCK-100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	98.1	225	5 AAE28837	Aae28837 Discosoma
2	1214	98.1	225	5 AAE17542	Aae17542 Discosoma
3	1204	97.3	225	5 AAE28836	Aae28836 Discosoma
4	1199	96.9	225	5 AAE28921	Aae28921 Discosoma
5	1199	96.9	225	5 AAE17541	Aae17541 Discosoma
6	1196	96.7	225	5 AAE28925	Aae28925 Discosoma
7	1195	96.6	225	5 AAE28919	Aae28919 Discosoma
8	1195	96.6	225	5 AAE28920	Aae28920 Discosoma
9	1195	96.6	225	7 ABW00937	Abw00937 Discosoma
10	1191	96.3	225	3 AAY99836	Aay99836 Discosoma

PD 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.
 XX 21-FEB-2001; 2001US-0270983P.
 PR 04-DEC-2001; 2001US-00006922.
 XX (CLON-) CLONTECH LAB INC.
 PA Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
 PI WPI; 2002-691654/74.
 XX N-PSDB; AAD46282.
 DR New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.
 XX
 XX Disclosure; Page: 80pp; English.
 PS The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX
 XX Sequence 225 AA;
 SQ

Alignment Scores:
 Pred. No.: 4,99e-102 Length: 225
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.14% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x AAE28837 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCCAGCTTCATCGCTTCAAGGTGGCGATGGAGGC 60
 DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACCGCCAGGTCGAGTCGAGGCGGAGGCGGCGCCCTACGAGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAACACCGCTGAAGTTGAAGGTGACCAAGGCGGCGCCCTCGCTTCCTGGGACATC 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCCGCCGAGTTCAGTACGCTCCAGGTGAGTGTAGTGAAGCACCCCGCCGACATPCCC 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAGAAGTGTCTTCCCGAGGCGTTCAGTGGGAGCGGTGATGAACCTTCGAG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGCGCGCTGGCGACCGGTGACCCAGGACTCTCTCCCTCAGAGCGGCTTCATCTAC 360
 DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGGTGAAGTTCATCGGCTGAACTTCCCTCCGAGCGGCGGCTGATGAGAGAGACC 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

QY 421 ATGGGCTGGAGGCTCCACCGAGCGCTGTACCCCGGAGCGGCTGCTGAAAGGCGAG 480
 DB 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCACAGGCGCTGAAGCTGAAGGACGCGCGCCACTACCTGGTGGAGTTCAAGTCCATC 540
 DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TCATGGCCAAAGACCCGCTGCAGCTGCCGCTACTACTACGTGGACACCAAGCTGGAC 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
 QY 601 ATCACCCTCCCAACAGGAGCTACACCATCTGGAGGAGTACGAGCGCAGCGGCGGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTTCTCTG 675
 DB 221 HisHisLeuPheLeu 225

RESULT 2
 AAE17542
 ID AAE17542 standard; protein; 225 AA.
 XX AAE17542;
 AC AAE17542;
 XX 22-APR-2002 (first entry)
 DT Disosoma sp. non-aggregating mutant timer protein ESNA.
 XX Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic plant; gene modification;
 KW protein age; mutant; ESNA; mutein.
 XX Disosoma sp.
 OS Synthetic.
 OS Key Location/Qualifiers
 FH Misc-difference 2 /note= "Wild type Arg substituted with Ala"
 FT Misc-difference 5 /note= "Wild type Lys substituted with Glu"
 FT Misc-difference 9 /note= "Wild type Lys substituted with Thr"
 FT Misc-difference 9 /note= "Wild type Lys substituted with Thr"
 XX WO200196373-A2.
 XX 20-DEC-2001.
 XX 13-JUN-2001; 2001WO-US019097.
 XX 14-JUN-2000; 2000US-0211607P.
 XX (CLON-) CLONTECH LAB INC.
 PA Fradkov AF, Tersikh A;
 PI WPI; 2002-154595/20.
 DR N-PSDB; AAD28209.
 XX New fluorescent timer proteins comprising an emission spectrum that
 PT changes over time from a first wavelength to a second wavelength, useful
 PT for monitoring intracellular protein movement, translocation, trafficking
 PT or stability.
 XX Claim 5; Fig 3; 89pp; English.
 PS The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in

CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 4,08e-101 Length: 225
 Score: 1204.00 Matches: 223
 Percent Similarity: 99.11% Conservativity: 0
 Best Local Similarity: 99.11% Mismatches: 2
 Query Match: 97.33% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x AAE28836 (1-225)

QY 1 ATGGCTCTCCCGAGAACGTCATCACGAGTTCATGGCTTCAAGTGGCGATGAGGGC 60
 DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGGCCACAGTTCGAGATCGAGGGCGAGGGCGGCCCGCCCTACGAGGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGGGGCCCGCTCCCTCCCTGGGACATC 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCCGCCAGTTCACAGTTCAGGTGAGTGAAGCACCCCGCCGACATCCCG 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAAAGAGTCTCTCCCGAGGGCTTCAAGTGGAGGGCGGTGATGAATCTCGAG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGGCGCTGGCGACCGGTGACCGAGGACTCTCCCTCGAGGACGGCTGCTCATCTAC 360
 DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGTGAAGTTCATCGCGGTGAATCTCCCTCGAGCGGCCCGCTGATGAGAGAGACC 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGGACGGCTGTGAAGGCGAG 480
 DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCAAAGCCCTGAAGTGAAGGACGGCGGCCACTACTGCTGAGTCAAGTCCATC 540
 DB 161 ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGCCCAAGAGCCGTCGAGTCCCGGCTACTACTAGTGCACACCCAGCTGGAC 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaLysLeuAsp 200
 QY 601 ATCACTCCCAACAGGAGTACTACCATCTGTGGAGCAGTACGAGCGCACCGGCGCGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACACCTGTTCCTG 675
 DB 221 HisHisLeuPheLeu 225

RESULT 4

ID AAE28921 standard; protein; 225 AA.

XX AAE28921;

AC
 XX

DT 27-DEC-2002 (first entry)
 XX Discosoma sp. drFP583 (NFP-6) mutant protein, E5 (V105A-S197T).
 DE Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 XX fluorescence activated cell sorting application; fluorescent timer;
 KW biosensor; fluorescence resonance energy transfer application; FRET;
 KW colouring agent; recombinant DNA application; analyte detection assay;
 KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
 KW mutin.
 XX Discosoma sp.
 OS Synthetic.
 OS
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 105 /note= "Wild-type Val substituted with Ala"
 FT Misc-difference 217 /note= "Wild-type Ser substituted with Thr"
 FT
 XX WO200268459-A2.
 PN
 XX 06-SEP-2002.
 PD
 XX 20-FEB-2002; 2002WO-US005749.
 PF
 XX 21-FEB-2001; 2001US-0270983P.
 PR 04-DEC-2001; 2001US-00006922.
 PR
 XX (CLON-) CLONTECH LAB INC.
 PA
 XX Lukanov S, Lukanov K, Yanushevich Y, Savitsky A, Fradkov A;
 PI WPI; 2002-691654/74.
 XX
 XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.
 PS Disclosure; Page; 80pp; English.
 XX
 CC The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 1.17e-100 Length: 225
 Score: 1199.00 Matches: 222
 Percent Similarity: 99.11% Conservativity: 1
 Best Local Similarity: 98.67% Mismatches: 2
 Query Match: 96.93% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x AAE28921 (1-225)

QY 1 ATGGCTCTCCCGAGAACGTCATCACGAGTTCATGGCTTCAAGTGGCGATGAGGGC 60
 DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGGCCCGCTACGAGGGC 120

Db 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGGTGAACCTCCCTCCGACGGCCCGCTGATGAGAAAGACG 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGACGGCGGTGCTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAGCTGAGGAGCGGCGCCACTACCTGCTGAGTTCAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCCAAGAGCGCGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
 Qy 601 ATCACTCCACAGAGGACTACACCATCTGCGAGCAGTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675
 Db 221 HisHisLeuPheLeu 225

RESULT 6
 AAE28925 standard; protein; 225 AA.
 XX AC AAE28925;
 XX DT 27-DEC-2002 (first entry)
 XX DE Discosoma sp. drFP583 (NFP-6) mutant protein, AG4 (V71M+V105A+S197T).
 XX KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 XX KW biosensor; fluorescence resonance energy transfer application; FRET;
 XX KW colouring agent; recombinant DNA application; analyte detection assay;
 XX KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
 XX KW mutin.
 XX OS Discosoma sp.
 XX OS Synthetic.
 XX PH Key Location/Qualifiers
 FT Misc-difference 71 /note= "Wild-type Val substituted with Met"
 FT FT Misc-difference 105 /note= "Wild-type Val substituted with Ala"
 FT FT Misc-difference 197 /note= "Wild-type Ser substituted with Thr"
 XX W0200268459-A2.
 XX 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.
 XX 21-FEB-2001; 2001US-0270983P.
 XX 04-DEC-2001; 2001US-00006922.
 XX (CLON-) CLONTECH LAB INC.
 XX Lukanov S, Lukanov K, Yanushevich Y, Savitsky A, Fradkov A;
 XX WPI; 2002-691654/74.
 XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 XX of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 XX analyte detection assays or fluorescence activated cell sorting
 XX applications.

XX PS Disclosure; Page; 80pp; English.
 CC The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX SQ Sequence 225 AA;
 Alignment Scores:
 Pred. No.: 2,19e-100 Length: 225
 Score: 1196.00 Matches: 221
 Percent Similarity: 99.11% Conservative: 2
 Best Local Similarity: 98.22% Mismatches: 2
 Query Match: 96.69% Indels: 0
 DB: 5 Gaps: 0
 US-10-081-864-14 (1-678) x AAE28925 (1-225)
 Qy 1 ATGCGCTCTCCGAGAACGTATCACCGAGTTCATCGCTTCAAGTGCATGAGGCG 60
 Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 Qy 61 ACCGTGAACGCGCAGGATTCGAGATCGAGGCGGCGAGGCGGCCCGCTACGAGGCG 120
 Db 21 ThrValAsnGlyHisGluPheGluIleGlyGlyGlyGlyGlyArgPheTyrGlyGly 40
 Qy 121 CACACACCGGTGAAGTGAAGGTGACCAAGGCGGCGGCCCTTCCGCTCGGACATC 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTyrAspIle 60
 Qy 181 CTGTCCCGCCGAGTTCAGTACGCTCAAGGTGTACGTGAAGCACCCCGCGGACATCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysMetTyrValLysHisProAlaAspIlePro 80
 Qy 241 GACTACAAGAGAGCTGTCTTCCCGAGGCGTCAAGTGGGAGCGCGTGTAGTGAACCTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 Qy 301 GACGCGCGGTGGCGACCGGTGACCCGAGACTCTTCCCTCGAGGACGCTGCTTCATCTAC 360
 Db 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGGTGAACCTCCCTCCGACGGCCCGCTGATGAGAAAGACG 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGACGGCGGTGCTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAGCTGAGGAGCGGCGCCACTACCTGCTGAGTTCAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCCAAGAGCGCGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
 Qy 601 ATCACTCCACAGAGGACTACACCATCTGCGAGCAGTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675

Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheLeuTrpAspIle 60
 Qy 181 CTGTCCTCCAGTTCCAGTACGGCTCCAGAGTGTAGTGAAGCACCCCGCCGACATCCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 Qy 241 GACTACAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGAGGCGCTGATGAACCTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrPduLargValMetAsnPheGlu 100
 Qy 301 GACGGCGGCTGGCGCGCTGACCGAGTCTCTCCCTGCGAGGCGCTGCTTCATCTAC 360
 Db 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGCTGAATCTCCCTCCGAGCGGCCCTGATGAGAGAGAGACC 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCGCTCCAGCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAGCTGAAGCTGAAGGAGCGGCGGCGCTACTCTGTTGAGTTCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCAAGAGCGCGCTGAGCTGCGAGTCTCTCCCTGCGAGGCGGCTGTGAGGCGAG 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 Qy 601 ATCACTCCACACAGGAGTACACCATCGTGAGCAGTACGAGCGGCGGCGGCGGCGGCG 660
 Db 201 IleThrSerHisGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675
 Db 221 HisHisLeuPheLeu 225

RESULT 11
 ID AAB01622 standard; protein; 225 AA.
 AC AAB01622;
 XX
 XX DT 12-DEC-2000 (first entry)
 XX DE Discosoma sp. red fluorescent protein drFP583.
 XX KW Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;
 XX XM zFP506; zFP538; drFP583; dsFP483; asFP600; dgFP512; dmFP592.
 XX OS Discosoma sp.
 XX PN WO200034526-A1.
 XX PD 15-JUN-2000.
 XX PF 10-DEC-1999; 99WO-US029405.
 XX PR 11-DEC-1998; 98US-00210330.
 XX PA (CLON-) CLONTECH LAB INC.
 XX PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
 XX DR WPI; 2000-423451/36.
 XX PT Novel method for identifying a DNA sequence encoding fluorescent proteins
 XX PT from non-bioluminescent Anthozoa which are useful for fluorescent
 XX PT labeling and as markers.
 XX PS Claim 3; Page 69-69; 73pp; English.
 XX

CC The present sequence is Discosoma sp. red fluorescent protein drFP583. It
 CC was isolated using the Aequoria victoria green fluorescent protein (GFP)
 CC sequence, which was used to design PCR primers which might isolate other
 CC fluorescent proteins from a number of species of Anthozoa. These were
 CC Anemonia majano, Clavularia sp., Zoanthus sp., Discosoma sp. and Anemonia
 CC sulcata. The cDNA obtained was then screened in the search for sequences
 CC encoding fluorescent proteins. The other proteins found in this manner
 CC were cFP484, zFP506, zFP538, amFP486, dsFP483, asFP600, dgFP512 and
 CC dmFP592. These proteins can be used as fluorescent labels (for gene
 CC expression and protein localisation studies) and in fluorescence resonance
 CC energy transfer (FRET) studies) in place of fluorophore derivatives and
 CC luciferases, as these involve laborious processes and the latter require
 CC cofactors. They can also be used in place of GFP, which is too stable to
 CC be useful when studying short-term or repetitive events
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 6.26e-100 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.57% Conservative: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 3 Gaps: 0

US-10-081-864-14 (1-678) x AAB01622 (1-225)

Qy 1 ATGGCTCTCCGAGAGCTCATCACCGAGTTCATGCGCTTCAAGTGCAGTGGAGGCG 60
 Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 Qy 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGCGGAGGCGGAGGCGGCGGCGGCGGCGG 120
 Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 Qy 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 Qy 181 CTGTCCCGCCAGTTCAGTACGGCTCCAGGTGTAGTGAAGCACCCCGCCGACATCCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 Qy 241 GACTACAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGAGGCGGCTGATGAACCTTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrPduLargValMetAsnPheGlu 100
 Qy 301 GACGGCGGCTGGCGAGCGCTGACCGAGTCTCTCCCTGCGAGGCGGCTGCTTCATCTAC 360
 Db 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGCTGAATCTCCCTCCGAGCGGCCCTGATGAGAGAGAGACC 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCGCTCCAGCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAGCTGAAGCTGAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCAAGAGCGCGCTGAGCTGCGAGTCTCTCCCTGCGAGGCGGCTGTGAGGCGAG 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 Qy 601 ATCACTCCACACAGGAGTACACCATCGTGAGCAGTACGAGCGGCGGCGGCGGCGGCGGCGG 660
 Db 201 IleThrSerHisGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675

Wed Aug 7 05:01:02 2002

Db	41	HisAsnThrValIysLeuIysValThrIysGlyGlyProLeuProPheAlaTrpAspIle	60
Qy	181	CTCTCCCCCAGTTCACAGTCCGCTCAACGGTGTACGTGAAGACCCCGCGGACATCCCC	240
Db	61	LeuSerProGlnPheGlnIyrGlySerLysValIyrVallysHisProAlaAspIlePro	80
Qy	241	GACTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGAATCAACTTCGAG	300
Db	81	AspIyrIysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Qy	301	GACGGCGGTGCGGACCGGTGACCCAGGACTCTCTCCCTGAGGACGGCTGCTTCATCTAC	360
Db	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleIyr	120
Qy	361	AAGGTGAAGTTCATCGCGTGAATCTCCCTCCGACGGCCCGGTGATGCAGAAAGACCC	420
Db	121	LysValIysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Qy	421	ATGGCTGGGAGGCCTCCACGAGCGCTGTACCCCGCGACGGCGTGTGAAGGGCGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuIyrProArgAspGlyValLeuLysGlyGlu	160
Qy	481	ATCCACAGGCCCTGAGCTGAAGGACGCGGGCCACTACCTGTGGGAGTTCAAGTCCATC	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisIyrLeuValGluPheLysSerIle	180
Qy	541	TACATGCCCAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACACCAAGCTGGAC	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyIyrIyrIyrValAspSerLysLeuAsp	200
Qy	601	ATCACCTCCACAACAGGAGGACTACACCATCGTGGAGCAGTACGACCGACCGAGGCCGC	660
Db	201	IleThrSerHisAsnGluAspIyrThrIleValGluGlnIyrGluArgThrGluGlyArg	220
Qy	661	CACCACCTGTTCTCTG 675	
Db	221	HisHisLeuPheLeu 225	
RESULT 13			
Db	ABB08834		
XX	ID	ABB08834 standard; protein; 225 AA.	
AC	AC	ABB08834;	
XX	XX	29-MAY-2002 (first entry)	
DT	XX	Yeast optimised RFP SEQ ID NO 17.	
DE	XX	Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;	
XX	KW	Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;	
KW	KW	Escherichia coli; green fluorescent protein; biotechnology.	
XX	OS	Anthozoa.	
XX	XX	DE20001395-U1.	
XX	XX	15-MAR-2001.	
XX	XX	27-JAN-2000; 2000DE-02001395.	
XX	XX	27-JAN-2000; 2000DE-02001395.	
XX	XX	(GPCB-) GPC BIOTECH AG.	
XX	XX	WPI; 2002-228394/29.	
DR	DR	N-FSDB; ABA93905, ABA95921, ABA95922.	
XX	XX	New DNA encoding red fluorescent protein, useful as marker in	
PT	PT	biotechnology, has sequence optimized for expression in eukaryotes,	
PT	PT	especially yeast or plants.	
XX	XX	Disclosure; Page 13-14; 19pp; German.	
PS	XX		

CC The invention relates to DNA (I) containing either sequence ABA95905 or
 CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
 CC (yRFP). (I) are used to express red fluorescent protein (RFP) in
 CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
 CC plants, especially dicotyledonous plants including Nicotiana tabacum or
 CC Arabidopsis thaliana and also in prokaryotes, especially bacteria
 CC especially Escherichia coli. RFP is useful in the same way as green
 CC fluorescent protein but is more generally applicable in modern
 CC biotechnology. (I) are optimised for expression in yeast and so generate
 CC RFP at higher levels with stronger fluorescence and thus lowers the
 CC detection limit and gives a better signal-to-noise ratio. The present
 CC sequence is that of the yeast optimised RFP
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 6.26e-100 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.57% Conservative: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x ABB08834 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATGGCTTCAAGGTGCGCATGGAGGC 60
 DB 1 MetArgSerSerLysAsnValLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGCGGCCGCCCTACGAGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluLeuGlyGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAACACCGTGAAGTTGAAGTGAACCAAGGGGGCCCCCTGCGCTTCGCTGGGACATC 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCTCCCGAGTTCAGGTCCAGGTGTCAGTGAAGCAGCCCGCCGACATCCCC 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAGAGCTGCTCTCTCCCGAGGGCTTCAAGTGGAGGCGCGTGAATCTCGAG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGCGCGCTGGCGACCGTGAACCGAGTCTCTCCCTGCGAGACGCGCTGCTCATCTAC 360
 DB 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGGTGAAGTTCATCGCGGTGAATCTCTCCCTGCGAGCGGCCCTGATGCAAGAGACC 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGCTGGGAGGCTCCACGAGCGCTGCTACCCCGAGCGCGCTGCTGAAGGCGGAG 480
 DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCAAAGCCCTGAAGTGAAGAGCGGCGGCACTACTGCTGGAGTTCAGTCCATC 540
 DB 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGCCCAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGCAC 600
 DB 181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATCACTCCCAACAGGAGTACTACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACACCTGTTCTCG 675
 DB 221 HisHisLeuPheLeu 225

RESULT 14

AAE28926
 ID AAE28926 standard; protein; 225 AA.

XX AAE28926;

DT 27-DEC-2002 (first entry)

XX Discosoma sp. drpP583 (NFP-6) mutant protein, (V71M+V105A+Y120H+S197T).
 DE Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 KW fluorescence activated cell sorting application; fluorescent timer;
 KW biosensor; fluorescence resonance energy transfer application; FRET;
 KW colouring agent; recombinant DNA application; analyte detection assay;
 KW sunscreen; second messenger detector; drpP583 protein; NFP-6; mutant;
 KW mutin.

XX Discosoma sp.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 71 /note= "Wild-type Val substituted with Met"
 FT Misc-difference 105 /note= "Wild-type Val substituted with Ala"
 FT Misc-difference 120 /note= "Wild-type Tyr substituted with His"
 FT Misc-difference 197 /note= "Wild-type Ser substituted with Thr"

PN WO200268459-A2.

XX 06-SEP-2002.

XX 20-FEB-2002; 2002WO-US005749.

XX 21-FEB-2001; 2001US-0270983P.

XX 04-DEC-2001; 2001US-00006922.

PA (CLON-) CLONTECH LAB INC.

XX Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Pradkov A;
 PI WPI; 2002-691654/74.

XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
 an analyte detection assays or fluorescence activated cell sorting
 applications.

XX Disclosure; Page: 80pp; English.

XX The invention relates to nucleic acid molecules encoding non-aggregating
 chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 useful in analyte detection assays, as colouring agents, as markers in
 recombinant DNA applications, as sunscreens or filters, in fluorescence
 resonance energy transfer (FRET) applications, as biosensors in
 prokaryotic and eukaryotic cells, in screening assays, as second
 messenger detectors, in fluorescence activated cell sorting applications,
 in protease cleavage assays or as fluorescent timers. The present
 sequence is Discosoma sp. drpP583 (NFP-6) mutant protein of the
 invention. Note: This sequence is not shown in the specification, but is
 derived from Discosoma sp. drpP583 (NFP-6) wild-type protein shown as SEQ
 ID NO:8 (AAE288833) in page 70-71 of the specification

XX Sequence 225 AA;

Alignment Scores:

Pred. No.: 6.26e-100 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 99.11% Conservative: 3
 Best Local Similarity: 97.78% Mismatches: 2
 Query Match: 96.28% Indels: 0

XX PF 20-FEB-2002; 2002WO-US005749.
XX PR 21-FEB-2001; 2001US-0270983P.
XX PR 04-DEC-2001; 2001US-00006922.
XX PA (CLON-) CLONTECH LAB INC.
XX XX Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX WPI; 2002-691654/74.
XX N-PSDB; AD46278.
XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX of an aggregating Chidarian chromo- or fluorescent protein or mutant for
XX analyte detection assays or fluorescence activated cell sorting
XX applications.
XX PS Disclosure; Page 70-71; 80pp; English.
XX CC The invention relates to nucleic acid molecules encoding non-aggregating
XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
XX useful in analyte detection assays, as colouring agents, as markers in
XX recombinant DNA applications, as sunscreens or filters, in fluorescence
XX resonance energy transfer (FRET) applications, as biosensors in
XX prokaryotic and eukaryotic cells, in screening assays, as second
XX messenger detectors, in fluorescence activated cell sorting applications,
XX in protease cleavage assays or as fluorescent timers. The present
XX sequence is Discosoma sp. drFP583 (NFP-6) wild-type protein of the
XX invention
XX SQ Sequence 225 AA;
Alignment Scores:
Pred. No.: 6.26e-100 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: Gaps: 5
US-10-081-864-14 (1-678) x AAE28833 (1-225)
QY 1 ATGGCTCTCTCCGAGAACGTCACTCCAGGTTTCATCGCTTCAAGGTGCGATGAGGCG 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCAGATCGAGGGCGAGGGCGCGCCCTTACGAGGCG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTCAGTCAAGGTGACCAAGGGCGGCCCTTCCCTTCCGTCGACATC 180
DB 41 HisAsnThrValLysLeuSerPheProGluGlyPheLysGlyGlyProLeuPheAlaTrpAspIle 60
QY 181 CTGTCTCCCGCCAGTTCAGTCCAGGCTCCAGGTGTCAGTGAAGCACCACCGCGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACACGAGCTGCTCTCCCGAGGCTTCAAGTGGGAGCGCGTGTATGATTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGACCGTTCAGTCCAGGACTCTCTCCCTCCAGGACGGTGTCTTCTATC 360
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGGCGTGAACCTTCCCTCCGAGCGCCCGTGTATGCAGAGAAGACC 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTTCCACCGAGCGCTGTATCCCGCGACCGCGGTGTGTGAAGCGCGAG 480

DB: 5 Gaps: 0
US-10-081-864-14 (1-678) x AAE28826 (1-225)
QY 1 ATGGCTCTCTCCGAGAACGTCACTCCAGGTTTCATCGCTTCAAGGTGCGATGAGGCG 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCAGATCGAGGGCGAGGGCGCGCCCTTACGAGGCG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTCAGTCAAGGTGACCAAGGGCGGCCCTTCCCTTCCGTCGACATC 180
DB 41 HisAsnThrValLysLeuSerPheProGluGlyPheLysGlyGlyProLeuPheAlaTrpAspIle 60
QY 181 CTGTCTCCCGCCAGTTCAGTCCAGGCTCCAGGTGTCAGTGAAGCACCACCGCGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysMetTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACACGAGCTGCTCTCCCGAGGCTTCAAGTGGGAGCGCGTGTATGATTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGACCGTTCAGTCCAGGACTCTCTCCCTCCAGGACGGTGTCTTCTATC 360
DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleHis 120
QY 361 AAGGTGAAGTTCATCGGCGTGAACCTTCCCTCCGAGCGCCCGTGTATGCAGAGAAGACC 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTTCCACCGAGCGCTGTATCCCGCGACCGCGGTGTGTGAAGCGCGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGGCGCTGAAGTTCAGGAGCGGCGGCGGCTCACTCTGCTGAGTTCAGTCCATC 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLysLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAACCGCGTCAAGTTCAGTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
QY 601 ATACCTCCCAACAGGAGCTACCATCTGTGAGCAGTACGAGCGCACCGAGGCGCGC 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCTG 675
DB 221 HisHisLeuPheLeu 225
RESULT 15
AAE28833
ID AAE28833 standard; protein; 225 AA.
AC AAE28833;
XX 27-DEC-2002 (first entry)
DT Discosoma sp. drFP583 (NFP-6) wild-type protein.
DE Fluorescent protein; chromoprotein; protease cleavage assay; filter;
XX fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunsreen; second messenger detector; drFP583 protein; NFP-6.
XX Discosoma sp.
OS WO200268459-A2.
XX 06-SEP-2002.
PD

Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
QY	481	ATCCACAAGGCCCTGAAGCTGAAGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
QY	541	TACATGGCCAAAGACCCGTCAGCTGCCCGCTACTACTACGTGGACACCAAGCTGGAC	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp	200
QY	601	ATCACCTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
QY	661	CACCACCTGTTCTCTG	675
Db	221	HisIleLeuPheLeu	225

Search completed: July 29, 2004, 14:32:26
Job time : 51.2188 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:29:01 ; Search time 13.9516 Seconds

(without alignments)
5017.682 Million cell updates/sec

Title: US-10-081-864-14

Perfect score: 1237

Sequence: 1 atggctctccgagacgt.....gccaccactgtctctgtaa 678

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_29072004_150507_25519/app_query.fasta_1.1742
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10081864 @CGN_1_1_45 @runat_29072004_150507_25519 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/prodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCUTS.COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	38.4	238	3	US-09-277-716-16
2	475	38.4	238	4	US-09-609-161B-16
3	475	38.4	238	4	US-09-626-581D-65
4	475	38.4	238	4	US-09-415-765B-65
5	475	38.4	238	4	US-09-626-580C-65
6	463	37.4	238	4	US-09-839-650-3
7	462	37.3	238	3	US-09-277-716-32
8	462	37.3	238	4	US-09-609-161B-32
9	229.5	18.6	238	4	US-09-023-946B-23
10	228.5	18.5	238	4	US-09-023-946B-28
11	226.5	18.3	238	2	US-08-818-604-32
12	226.5	18.3	238	3	US-08-819-612-22

13	226.5	18.3	238	4	US-09-316-319-3	Sequence 3, Appl
14	226.5	18.3	238	4	US-09-346-946-32	Sequence 32, Appl
15	226.5	18.3	238	4	US-09-023-946B-22	Sequence 22, Appl
16	224.5	18.1	238	1	US-08-337-915A-2	Sequence 2, Appl
17	224.5	18.1	238	1	US-08-753-143-2	Sequence 2, Appl
18	224.5	18.1	238	2	US-08-679-865-2	Sequence 2, Appl
19	224.5	18.1	238	2	US-08-680-876-2	Sequence 2, Appl
20	224.5	18.1	238	2	US-08-792-553-2	Sequence 2, Appl
21	224.5	18.1	238	3	US-08-911-825-2	Sequence 2, Appl
22	224.5	18.1	238	3	US-08-753-144-2	Sequence 2, Appl
23	224.5	18.1	238	3	US-08-974-737-2	Sequence 2, Appl
24	224.5	18.1	238	3	US-08-706-408-2	Sequence 2, Appl
25	224.5	18.1	238	3	US-09-094-359-2	Sequence 2, Appl
26	224.5	18.1	238	3	US-09-172-063-2	Sequence 1, Appl
27	224.5	18.1	238	3	US-09-121-539-1	Sequence 1, Appl
28	224.5	18.1	238	3	US-09-263-975-2	Sequence 2, Appl
29	224.5	18.1	238	4	US-08-737-452-2	Sequence 2, Appl
30	224.5	18.1	238	4	US-09-465-142-2	Sequence 2, Appl
31	224.5	18.1	238	4	US-09-418-785-1	Sequence 1, Appl
32	224.5	18.1	238	4	US-09-214-509-2	Sequence 2, Appl
33	224.5	18.1	238	4	US-09-479-645A-10	Sequence 159, App
34	224.5	18.1	238	4	US-09-479-645A-159	Sequence 2, Appl
35	224.5	18.1	238	4	US-09-129-192C-2	Sequence 74, Appl
36	224.5	18.1	238	4	US-09-129-192C-74	Sequence 2, Appl
37	224.5	18.1	238	4	US-09-575-847-2	Sequence 2, Appl
38	224.5	18.1	238	4	US-09-602-641-2	Sequence 2, Appl
39	224.5	18.1	238	4	US-09-704-463-2	Sequence 2, Appl
40	224.5	18.1	238	4	US-09-472-065A-4	Sequence 4, Appl
41	224.5	18.1	238	4	US-09-603-448-7	Sequence 7, Appl
42	224.5	18.1	238	4	US-09-603-448-24	Sequence 24, Appl
43	224.5	18.1	238	4	US-09-023-946B-24	Sequence 24, Appl
44	224.5	18.1	238	4	US-09-920-922-4	Sequence 2, Appl
45	224.5	18.1	238	5	PCT-US95-14692-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-277-716-16

; Sequence 16, Application US/09277716A

; Patent No. 6232107

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

; CURRENT APPLICATION NUMBER: US/09/277,716A

; CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: 60/102,939

; EARLIER FILING DATE: 1998-10-01

; EARLIER APPLICATION NUMBER: 60/089,367

; EARLIER FILING DATE: 1998-06-15

; EARLIER APPLICATION NUMBER: 60/079,624

; EARLIER FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 16

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Renilla mulleri

; FEATURE:

; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Alignment Scores:

Pred. No.: 2, 24e-34 Length: 238

Score: 475.00 Matches: 93

Percent Similarity: 64.59% Conservative: 42

Best Local Similarity: 44.50% Mismatches: 72

Query Match: 38.40% Indels: 2

DB: 3 Gaps: 2

US-10-081-864-14 (1-678) x US-09-277-716-16 (1-238)

```

QY 22 ATCCAGGTTTCATCGCTCAAGTGGCGATGGAGGCGACCGTGAACGGCCACGAGTTC 81
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnAsnHisValPhe 30
QY 82 GAGATCGAGGCGGAGGCGGCGCCCTACAGGGCCACCAACACCGTGAAGTTGAAG 141
DB 31 ThrMetGluGlyCysGlyLysGlyAsnLeuPheGlyAsnGlnLeuValGlnLeuArg 50
QY 142 GTGACCAAGGGCGGCGCCCTCGCTCGCTGGACATCTCTCCCGGAGTTCAGTAC 201
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspLeuValSerProAlaPheGlnTyr 70
QY 202 GGCTCCAAAGTGTACGTGAAGACCCCGCGACATCTCTCAAGGTGAAGTTCATCGCTTC 261
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspLeuValSerProAlaPheGlnTyr 90
QY 442 GAGCGCTGTACCCCGCGAGCGGTCTGATGAGGCGGAGATCCACAGGCGCTGAAGCTG 501
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluLeuValTyrLysLeu 170
QY 502 AAGGACGGCGCCACTACCTGGTGGAGTTCAGTTCATCTACATGCGCCAGAGCCCGCTG 561
DB 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
QY 562 ---CAGTGGCGGCTACTACTACCTGGAGTTCAGTTCATCTACATGCGCCAGAGCCCGCTG 618
DB 191 LysGluPheProSerTyrHisPheLeuGlnHisArgLeuGluLysThr---TyrValGlu 209
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
DB 210 AspGlyGlyPheValGluGlnHisGlu 218

RESULT 2
US-09-609-161B-16
; Sequence 16, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri

```

```

; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-16

Alignment Scores:
Pred. No.: 2,24e-34 Length: 238
Score: 475.00 Matches: 93
Percent Similarity: 64.59% Conservative: 42
Best Local Similarity: 44.50% Mismatches: 72
Query Match: 38.40% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-609-161B-16 (1-238)
QY 22 ATCCAGGTTTCATCGCTCAAGTGGCGATGGAGGCGACCGTGAACGGCCACGAGTTC 81
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnAsnHisValPhe 30
QY 82 GAGATCGAGGCGGAGGCGGCGCCCTACAGGGCCACCAACACCGTGAAGTTGAAG 141
DB 31 ThrMetGluGlyCysGlyLysGlyAsnLeuPheGlyAsnLeuValGlnLeuArg 50
QY 142 GTGACCAAGGCGGCGCCCTCGCTTCGCTGGACATCTCTCCCGGAGTTCAGTAC 201
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspLeuValSerProAlaPheGlnTyr 70
QY 202 GGCTCCAAAGTGTACGTGAAGACCCCGCGACATCTCTCAAGGTGAAGTTCATCGCTTC 261
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspLeuValSerProAlaPheGlnTyr 90
QY 262 CCCGAGGCTTCAAGTGGGCGGTGATGAGTTCGAGGCGGCGGTGGCGGCGCGTGC 321
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluLe 110
QY 322 ACCCAGGACTCCTCCCTCGAGGCGGTCTTCTACATCTCAAGGTGAAGTTCATCGCTTC 381
DB 111 ArgSerAspLeuAsnLeuLeuGluAspLysPheValTyrArgValGluTyrLysGlySer 130
QY 382 AACTTCCCTCCGAGCGCCCGGTGATGACGAAGACCATGGCTGGGAGCGCTCCACC 441
DB 131 AsnPheProAspAspGlyProValMetGlnLysThrLeuGlyLeuGluProSerPhe 150
QY 442 GAGCGCTGTACCCCGCGAGCGGTCTGATGAGGCGGAGATCCACAGGCGCTGAAGCTG 501
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluLeuValTyrLysLeu 170
QY 502 AAGGACGGCGCCACTACCTGGTGGAGTTCAGTTCATCTACATGCGCCAGAGCCCGCTG 561
DB 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
QY 562 ---CAGTGGCGGCTACTACTACCTGGAGTTCAGTTCATCTACATGCGCCAGAGCCCGCTG 618
DB 191 LysGluPheProSerTyrHisPheLeuGlnHisArgLeuGluLysThr---TyrValGlu 209
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
DB 210 AspGlyGlyPheValGluGlnHisGlu 218

RESULT 3
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08

```



```

; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PR1
; ORGANISM: Renilla muelleri
US-09-626-581D-65

```

Alignment Scores:	2.24e-34	Length:	238
Pred. No.:		Matches:	93
Score:	475.00	Conservative:	42
Percent Similarity:	64.59%	Mismatches:	72
Best Local Similarity:	64.50%	Indels:	2
Query Match:	38.40%	Gaps:	2
DB:	4		

US-10-081-864-14 (1-678) x US-09-626-581D-65 (1-238)

22	ATCACCGAGTTTCATCGCTTCAAGGTGGCCATGAGGGCACCGTGAACGGCCACGAGTTC	81
QY	...	
11	LeuGlnGluValMetSerTyrLysValAsnLeuGlyLeuValAsnAsnHisValPhe	30
Db	...	
82	GAGATCGAGGGGAGGGGAGGGCGGCCCTTACGAGGGCCACACACCGTGAAGTTGAAG	141
QY	...	
31	ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleAsz	50
Db	...	
142	GTGACCAAGGGCGGCCCTCGCTTCGGCTGGGACATCTGTCCCCCAAGTTCACGATAC	201
QY	...	
51	ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr	70
Db	...	
202	GGCTCCAAAGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAAGACTGTCTCTTC	261
QY	...	
71	GlyAsnArgThrPheThrLysTyr-ProAsnAspIleSerAspTyrPheIleGlnSerPhe	90
Db	...	
262	CCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGGACGGCGCGTGGCGACCGT	321
QY	...	
91	ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyLeuValGluIle	110
Db	...	
322	ACCCAGGACTCTCCCTCGAGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGCGCGT	381
QY	...	
111	A-gSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer	130
Db	...	
382	AACTTCCCTCCGACGGCCCGCTGATGCAGAAAGAACCATGGCTGGGAGGCGCTCCAC	441
QY	...	
131	AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe	150
Db	...	
442	GAGCGCTGTACCCCGCGACGGCGTGTGAAGGGCGAGATCCAAAGGCCCTGAAGCTG	501
QY	...	
151	GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValIleLeuValTyrLysLeu	170
Db	...	
502	AAGGACGGCGGCCTACTCTGTGGAGTTCAAGTCCATCTACATGGCCCAAGAACCCCGT	561
QY	...	
171	AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal	190
Db	...	
562	---CAGTCCCGGTACTACTACGTGGACACCAAGCTGGACATCACTCCCAACAGAG	618
QY	...	
191	LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu	209
Db	...	
619	GACTACACCATCGTGGAGCAGTACGAG	645
QY	...	
210	AspGlyGlyPheValGluGlnHisGlu	218
Db	...	

RESULT 4

US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6549632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold
; TITLE OF INVENTION: Libraries
; FILE REFERENCES: A66900-1/RWS/AMS
; CURRENT APPLICATION NUMBER: US/09/415-
; CURRENT FILING DATE: 1999-10-08

```

; PRIOR APPLICATION NUMBER: 09/169,015
;
; PRIOR FILING DATE: 1998-10-08
;
; NUMBER OF SEQ ID NOS: 65
;
; SOFTWARE: PatentIn ver. 2.0
;
; SEQ ID NO 65
;
; LENGTH: 238
;
; TYPE: PRT
;
; ORGANISM: Renilla muelleri
US-09-415-765B-65

```

Alignment Scores:	2.24e-34	Length:	238
Pred. No.:	475.00	Matches:	93
Percent Similarity:	64.50%	Conservative:	42
Best Local Similarity:	64.50%	Mismatches:	72
Query Match:	38.40%	Indels:	2
DB:	4	Gaps:	2

US-10-081-864-14 (1-678) x US-09-415-765B-65 (1-238)

Qy	22	ATCACCGAGTTCTATCGCTTCAGGTGCGCATGAGGGCACCGTGAACGGCCACGAGTTC	81
Db	11	LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnAsnHisValPhe	30
Qy	82	GAGATCGAGGGCGAGGGCGAGGGCGGCCCTACGAGGGCCACAAACCGCTGAAGTTGAAG	141
Db	31	ThrMetGluGlyCysGlyLysGlyAsnIleuPheGlyAsnGlnLeuValGlnIleArg	50
Qy	142	GTGACACAGGGCGGCCCTCCCTTCGCTGGGACATCTCTGTCCCGCCAGTTCCAGTAC	201
Db	51	ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr	70
Qy	202	GGCTCCACAGGTGTACGTGAAGCACCCCGCCGACATCCCGACTACAAAGAGCTGCTCTTC	261
Db	71	GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe	90
Qy	262	CCCGAGGGCTTCAGTGGGAGCGGTGATGAATCTGGAGACCGCGCTGGCGACCGTG	321
Db	91	ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle	110
Qy	322	ACCCAGGACTCTCCCTCAGAGCGGTGCTTCATCTACAAGGTGAAGTTCATCGCGGTG	381
Db	111	ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGlnTyrLysGlySer	130
Qy	382	AACTCTCCCTCCGACGGCCCGTGTATGACAGAAGACCATGGGTGGAGGCGCTCCAC	441
Db	131	AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe	150
Qy	442	GAGCGCTGTATCCCCCGCAGCGGTGCTGAGGGCGAGATCCACAAGGCCCTGAGCTGTG	501
Db	151	GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValIleLeuValTyrLysLeu	170
Qy	502	AAGGACGGCGGCACCTACTCGTGGAGTTCAAAGTCCATCTACATGGCCCAAGACCGCGTG	561
Db	171	AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal	190
Qy	562	---CAGTCCCGCGGTACTACTACTAGTGGACACCAAGCTGCACATCACTCCCAACAGAG	618
Db	191	LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu	209
Qy	619	GACTACACCATCGTGAGCAGTACGAG	645
Db	210	AspGlyGlyPheValGlnGlnHisGlu	218

RESULT 5

US-03-626-580C-65
; Sequence 65, Application US/09626580C
; Patent No. 6563617
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD

FILE REFERENCE: A-66900-2/RMS/AMS
CURRENT APPLICATION NUMBER: US/09/626,580C
CURRENT FILING DATE: 2000-07-27
PRIORITY APPLICATION NUMBER: US 09/415,765
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 09/169,015
PRIOR FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO 65
LENGTH: 238
TYPE: PRT
ORGANISM: Renilla muelleri
US-09-626-580C-65

Alignment Scores:
Pred. No.: 2,24e-34 Length: 238
Score: 475.00 Matches: 93
Percent Similarity: 64.59% Conservative: 42
Best Local Similarity: 44.50% Mismatches: 72
Query Match: 38.40% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-626-580C-65 (1-238)

```
QY 22 ATACCGAGTTCATCGCTTCAGGTCCGATCGAGGCGCACCGTGAACGCCACGAGTTC 81
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnAsnHisValPhe 30
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 GAGATCAGGCGGAGGCGGAGGCGGCGCCCTACGAGGCGCACACACCGTGAAGTTGAAG 141
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 GTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 201
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 GGTCTCAAGGTGTACGTGAAGCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 261
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 262 CCCGAGGCGTTCAGTGGGAGCGCGGTGATGAATTCGAGACGCGCGCGCGCGCGCGG 321
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 322 ACCGAGGACTCCTCCCTCGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 381
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer 130
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 382 AACTTCCCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 441
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 442 GAGGCGCTGTACCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 501
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyValIleLeuValTyrLysLeu 170
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 AAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 561
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 562 ---CAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 618
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu 209
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 AspGlyGlyPheValGluGlnHisGlu 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6

US-09-626-580C-65
Sequence 3, Application US/09839650
Patent No. 6645761

GENERAL INFORMATION:
APPLICANT: Stratagene
TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
Patent No. 6645761
TITLE OF INVENTION: Fluorescent Protein
FILE REFERENCE: 25436/1755
CURRENT APPLICATION NUMBER: US/09/839,650
CURRENT FILING DATE: 2001-04-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 238
TYPE: PRT
ORGANISM: Renilla muelleri
US-09-839-650-3

Alignment Scores:
Pred. No.: 2,63e-33 Length: 238
Score: 463.00 Matches: 92
Percent Similarity: 63.64% Conservative: 41
Best Local Similarity: 44.02% Mismatches: 74
Query Match: 37.43% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-839-650-3 (1-238)

```
QY 22 ATACCGAGTTCATCGCTTCAGGTCCGATCGAGGCGCACCGTGAACGCCACGAGTTC 81
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnAsnHisValPhe 30
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 GAGATCAGGCGGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 141
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 ThrMetGluGlyCysArgLysArgAsnIleLeuPheGlyAsnGlnLeuValHisIleArg 50
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 GTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 201
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 GGTCTCAAGGTGTACGTGAAGCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 261
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 262 CCCGAGGCGTTCAGTGGGAGCGCGGTGATGAATTCGAGACGCGCGCGCGCGCGCGG 321
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 322 ACCGAGGACTCCTCCCTCGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 381
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer 130
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 382 AACTTCCCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 441
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 442 GAGGCGCTGTACCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 501
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyValIleLeuValTyrLysLeu 170
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 AAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 561
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 562 ---CAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 618
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu 209
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 AspGlyGlyPheValGluGlnHisGlu 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7

US-09-277-716-32
Sequence 32, Application US/09277716A

```

; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUNE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-277-716-32

Alignment Scores:
Pred. No.: 3,23e-33 Length: 238
Score: 462.00 Matches: 93
Percent Similarity: 63.64% Conservative: 70
Best Local Similarity: 44.50% Mismatches: 74
Query Match: 37.35% Indels: 2
DB: 3 Gaps: 2

US-10-081-864-14 (1-578) x US-09-277-716-32 (1-238)
QY 22 ATCACCAGGTCATCGCTTCAAGTGGCATGGAGGCGACCGTGAACGGCCACGAGTTC 81
DB :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 LeuLysGluLeuMetSerAlaLysAlaSerValGluGlyIleValAsnHisValPhe 30
QY 82 GAGATCAGGGCGAGGGCGAGGGCGGCCCTACGAGGGCCACACACCGCTGAAGTTGAAG 141
DB :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
31 SerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeuMetGlnIlearg 50
QY 142 GTGACCAAGGGCGGCCCTGCCCTTCCTGGGACATCTCTGCCCGGAGTTCAGTAC 201
DB 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIleAlaPheGlnTyr 70
QY 202 GGCTCCAAAGGTGACGTGAAGCACCCCGCGACATCCCGGACTACAAGAGCTGTCCTTC 261
DB 71 GlyAsnArgThrPheThrLysTyrProAspAspIleAlaAspTyrPheValGlnSerPhe 90
QY 262 CCCGAGGCTTCAAGTGGAGCGCGTGTATGAACCTTCGAGGACGGCGGCGTGGCAGCCGTG 321
DB 91 ProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAlaIleValAspIle 110
QY 322 ACCCAGGACTCTCCCTGCAGGACGGCTGCTTCTACAGGTGAAGTTCATCGCGGTG 381
DB 111 ArgSerAspIleSerLeuGluAspLysPheHisTyrLysValGluTyrArgGlyAsn 130
QY 382 AACCTTCCCTCCGAGCGGCCGTGATGCAGAGAAGACCATCGGTGGGAGCGCTCCACCC 441
DB 131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150
QY 442 GAGCGCTGTACCCCGCGACGGCGTGTGAAGGGCGAGATCCACAGGCCCTGAAGCTG 501
DB 151 GluValValTyrMetAsnSerGlyValLeuValGlyValGluValAspLeuValTyrLysLeu 170
QY 502 AAGCAGCGCGCCACTACTCGTGGAGTTCAGTCCATCTCAAGTCCATCATGCCAAGAAGCCCGTG 561
DB 171 GluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSerLysGlyVal 190
QY 562 ---CAGCTGCCCGGCTACTACTAGCTGGACACCAAGCTGGACATCACTCCCAACAGCG 618
DB 191 LysGluPheProGluTyrHisPheIleHisArgLeuGluLysThr---TyrValGlu 209

```

QY 619 GACTACACCATCTGGAGCAGTACGAG 645
:: :: |||||:::|
Db 210 GluGlySerPheValGluGlnHisGlu 218
::: :: |||||:::|

RESULT 8
US-09-609-161B-32
; Sequence 32, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-609-161B-32

Alignment Scores:
Pred. No.: 3, 23e-33 Length: 238
Score: 462.00 Matches: 93
Percent Similarity: 63.84% Conservative: 40
Best Local Similarity: 44.50% Mismatches: 74
Query Match: 37.35% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-609-161B-32 (1-238)

QY 22 ATCACCGAGTTCATGGCTTCAAGTGGCATGGAGGCGACCGTGACGGCCACGAGTTC 81
::: |||||:::|
Db 11 LeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsnAsnHisValPhe 30
::: |||||:::|
QY 82 GAGATCAGAGGCGAGGCGAGGCGCGCCCTACGAGGCGCACACACCGCTGAAGTTGAAG 141
::: |||||:::|
Db 31 SerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeuMetGlnIleArg 50
::: |||||:::|
QY 142 GTGACCAAGGGCGGCGCCCTCGCTTCGCTGGGACATCCTGTCCTCCCGAGTCCAGTAC 201
::: |||||:::|
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIleAlaPheGlnTyr 70
::: |||||:::|
QY 202 GGCTCCAAAGGTGTAGTGAAGACACCCCGCGACATCCCGACTACAAAGAAGCTGTCTTC 261
::: |||||:::|
Db 71 GlyAsnArgThrPheThrLysTyrProAspAspIleAlaAspTyrPheValGlnSerPhe 90
::: |||||:::|
QY 262 CCGAGGGCTTCATGTGGGAGCGCGGTGTATGAACTTCGAGGACGCGCGCGTGGCGACCGTG 321
::: |||||:::|
Db 91 ProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAlaIleValAspIle 110
::: |||||:::|
QY 322 ACCCAGGACTCCTCCCTCAGGACGGCTGCTTCATCTACAAGGTGAAGTTTCATCGGCGTG 381
::: |||||:::|
Db 111 ArgSerAspIleSerLeuGluAspAspLysPheHisTyrLysValGlnTyrArgGlyAsn 130
::: |||||:::|
QY 382 AACTTCCCTTCGAGGCGCGCGGTGTATGACGAAGAAGACCATCGGCTGGGAGGCTTCACC 441
::: |||||:::|

Db 131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150
 Qy 442 GAGCGCTGTACCCCGGACGGCTGCTGAAGGGGAGATCCACAAAGGCCCTGAAGCTG 501
 Db 151 GluValValTyrMetAsnSerGlyValLeuValGluValAspLeuValTyrLysLeu 170
 Qy 502 AAGGACGGCGCCACTACCTGCTGAGTTCAAGTCCATCATCATGCGCCAGAGCCCGTG 561
 Db 171 GluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSerLysGlyGlyVal 190
 Qy 562 ---CAGTGGCCGCTACTACTAGTGACACCAAGCTGCATCATCTACATGCCAGAGCCGAG 618
 Db 191 LysGluPheProGluTyrHisPheIleHisArgLeuGluLysThr---TyrValGlu 209
 Qy 619 GACTACACCATCGTGGAGCAGTACGAG 645
 Db 210 GluGlySerPheValGluGlnHisGlu 218

RESULT 9

US-09-023-946B-23
 ; Sequence 23 Application US/09023946B
 ; Patent No. 6670449

GENERAL INFORMATION:

APPLICANT: GERO MIESENBOCK, ET AL.
 TITLE OF INVENTION: HYBRID MOLECULES AND
 THEIR USE FOR OPTICALLY DETECTING CHANGES IN

CELLULAR MICROENVIRONMENTS

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,946B
 FILING DATE: 13-FEB-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/036,805
 FILING DATE: 14 FEBRUARY 1997
 APPLICATION NUMBER: 60/038,179
 FILING DATE: 13 FEBRUARY 1997

ATTORNEY/AGENT INFORMATION:

NAME: KENNETH H. SONNENFELD
 REGISTRATION NUMBER: 33,285

REFERENCE/DOCKET NUMBER: 2955-4004US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 TELE: 421792

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 238

TYPE: AMINO ACID

STRANDEDNESS: UNKNOWN

TOPOLOGY: UNKNOWN

SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-09-023-946B-23

Alignment Scores:

Pred. No.:	Score:	Length:	Mismatches:
1.76e-12	229.50	238	60
Percent Similarity:	47.21%	Conservative:	50
Best Local Similarity:	25.75%	Mismatches:	110
Query Match:	18.55%	Indels:	13
DB:	4	Gaps:	6

US-10-081-864-14 (1-678) x US-09-023-946B-23 (1-238)
 Qy 1 ATGGCTCTCCGAGAACGTTCATCACCAGATTTCATGCGCTTCAAGTGGCCATGAGGGC 60
 Db 1 MetSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
 Qy 61 ACCGTGAACGCCCGAGTTCGATCGAGGCGGAGGCGGAGGCGCGCCCTACGAGGCG 120
 Db 21 AspValAsnGlyHisLysPheSerValSerGlyGluGlyAspAlaThrTyrGly 40
 Qy 121 CACAACACCGTGAAGTTGAAGTGACCAAGGCGGCGCCCTGCGCTTCGCTGGGACATC 180
 Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProTyrProThr 59
 Qy 181 CTGTCCCGCCAGTTCAGTCCAGTCCAGGTGTACGTGAAGCACCCCGCGCATCCCC 240
 Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
 Qy 241 -----GACTACAAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAAC 294
 Db 80 ArgHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluAsgThrIlePhe 99
 Qy 295 TTCGAGGACGGCGGCGTGGCGACCGTGCACCCAGGACTCCTCCCTGCAGGACGGCTGCTC 354
 Db 100 PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
 Qy 355 ATCTACAAAGGTGAAGTTTCATCGCGTGAAGTTCCTCCCGAGGCGCGCGTGCATCAGAAG 414
 Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
 Qy 415 AAGACCATGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGCGAGGCGGCTGCTGAAG 474
 Db 140 Lys---LeuGluTyrAsnTyrAsnAspHisGlnValTyrIleMetAlaAspLysGlnLys 158
 Qy 475 GCGGAGATCCACAAGGCCCTGAAGCTGAAG-----GACGGCGGCGCACTACTCTG 522
 Db 159 AsnGlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlyGlyValGlnLeu 178
 Qy 523 GTGAGGTTCT-----AAGTCCATCTACATGCCCAAGAGCCCGTGCAGCTGCCCGGCTAC 576
 Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198
 Qy 577 TACTACGTGGACACCAAGCTGGACATCACC-----TCCCACACAGGAGGATACACC 627
 Db 199 HisTyrLeuHisThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMet 218
 Qy 628 ATCGTGGAGCAGTACGAGCGCACCGAGGCGCGGCCACCAC 666
 Db 219 ValLeuLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 10

US-09-023-946B-28

; Sequence 28, Application US/09023946B

; Patent No. 6670449

GENERAL INFORMATION:

APPLICANT: GERO MIESENBOCK, ET AL.

TITLE OF INVENTION: HYBRID MOLECULES AND

THEIR USE FOR OPTICALLY DETECTING CHANGES IN

CELLULAR MICROENVIRONMENTS

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:

```
/ APPLICATION NUMBER: US/09/023,946B
/ FILING DATE: 13-Feb-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/036,805
/ FILING DATE: 14 FEBRUARY 1997
/ APPLICATION NUMBER: 60/038,179
/ FILING DATE: 13 FEBRUARY 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KENNETH H. SONNENFELD
/ REGISTRATION NUMBER: 33,285
/ REFERENCE/DOCKET NUMBER: 2955-4004US2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ TELEX: 421792
/ INFORMATION FOR SEQ ID NO: 28:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 238
/ TYPE: AMINO ACID
/ STRANDEDNESS: UNKNOWN
/ TOPOLOGY: UNKNOWN
/ SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-023-946B-28

Alignment Scores:
Pred. No.: 2,17e-12 Length: 238
Score: 226.50 Matches: 60
Percent Similarity: 47.21% Conservative: 50
Best Local Similarity: 25.75% Mismatches: 110
Query Match: 18.47% Indels: 13
DB: 4 Gaps: 6

US-10-081-864-14 (1-678) x US-09-023-946B-28 (1-238)
QY 1 ATGGCTCTCCGAGAACGTCATCAGGATTCATCGGTTCAAGTTCGATGAGGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGTCGAGGGCGAGGGCGGCGCCCTACGAGGC 120
Db 21 AspValAsnGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAACGAGGGCGGCGCCCTTCGCTTCCGCTGGGACATC 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGATCGGCTCAAGGTTCAGTGAAGCACCCTCCGACATCCCC 240
Db 60 LeuValThrThrPheSerTyGlyValGlnCysPheSerArgTyProAspHisMetLys 79
QY 241 -----GACTACAAGAGCTGCTCTCCCGAGCGCTTCAAGTGGAGCGCGTGATGAAC 294
Db 80 ArgHisAspPhePheLysSerAlaMetProGluGlyTyValGlnGluArgThrIlePhe 99
QY 295 TTCGAGACGGCGGTGGCGACCGTGACCCGAGACTCTCTCCCTGAGGACGGCTGCTTC 354
Db 100 PheLysAspGlyAsnTyLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAGTGAAGTTCATCGGCGTCACTCCCTCCGAGCGCCCGGTGCTGAAG 414
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspLysIleLeuGlyHis 139
QY 415 AAGACCATGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGAGCGCGGTGCTGAAG 474
Db 140 Lys---LeuGluTyAsnTyAsnAspHisGlnValTyIleMetAlaAspLysGlnLys 158
QY 475 GCGAGATCCACAGGCCCTCGAAGCTGAAG-----CAACGGCGCCACTACCTG 522
Db 159 AsnGlyIleLysValAsnPheLysIleArgHisAsnIleGluAspGlyGlyValGlnLeu 178
QY 523 GTGAGATC-----AAGTCCATCTACATGCGCAAGAAGCCGCTGAGCTGCGCGCTAC 576
Db 179 AlaAspHisTyGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198

US-10-081-864-14 (1-678) x US-08-818-604-32 (1-238)
QY 1 ATGGCTCTCCGAGAACGTCATCAGGATTCATCGGTTCAAGTTCGATGAGGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGTCGAGGGCGAGGGCGGCGCCCTACGAGGC 120
Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGluGlyGluGlyAspAlaThr-TyrGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAACGAGGGCGGCGCCCTTCGCTTCCGCTGGGACATC 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGATCGGCTCAAGGTTCAGTGAAGCACCCTCCGACATCCCC 240
Db 60 LeuValThrThrPheSerTyGlyValGlnCysPheSerArgTyProAspHisMetLys 79
QY 241 -----GACTACAAGAGCTGCTCTCCCGAGCGCTTCAAGTGGAGCGCGTGATGAAC 294
Db 80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyValGlnGluArgThrIlePhe 99
QY 295 TTCGAGACGGCGGTGGCGACCGTGACCCGAGACTCTCTCCCTGAGGACGGCTGCTTC 354
Db 100 TyrLysAspGlyAsnTyLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAGTGAAGTTCATCGGCGTGAATTCGCTCCGAGCGCCCGGTGCTGAAG 414
Db 355 ATCTACAAGTGAAGTTCATCGGCGTGAATTCGCTCCGAGCGCCCGGTGCTGAAG 414

Alignment Scores:
Pred. No.: 3,27e-12 Length: 238
Score: 226.50 Matches: 60
Percent Similarity: 47.44% Conservative: 51
Best Local Similarity: 25.84% Mismatches: 108
Query Match: 18.31% Indels: 15
DB: 2 Gaps: 7

US-10-081-864-14 (1-678) x US-08-818-604-32 (1-238)
QY 1 ATGGCTCTCCGAGAACGTCATCAGGATTCATCGGTTCAAGTTCGATGAGGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGTCGAGGGCGAGGGCGGCGCCCTACGAGGC 120
Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGluGlyGluGlyAspAlaThr-TyrGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAACGAGGGCGGCGCCCTTCGCTTCCGCTGGGACATC 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGATCGGCTCAAGGTTCAGTGAAGCACCCTCCGACATCCCC 240
Db 60 LeuValThrThrPheSerTyGlyValGlnCysPheSerArgTyProAspHisMetLys 79
QY 241 -----GACTACAAGAGCTGCTCTCCCGAGCGCTTCAAGTGGAGCGCGTGATGAAC 294
Db 80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyValGlnGluArgThrIlePhe 99
QY 295 TTCGAGACGGCGGTGGCGACCGTGACCCGAGACTCTCTCCCTGAGGACGGCTGCTTC 354
Db 100 TyrLysAspGlyAsnTyLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAGTGAAGTTCATCGGCGTGAATTCGCTCCGAGCGCCCGGTGCTGAAG 414
Db 355 ATCTACAAGTGAAGTTCATCGGCGTGAATTCGCTCCGAGCGCCCGGTGCTGAAG 414
```

Db 120 ValAsnArgIleGluLeuLysGlyLeuAspPheLysGluAspGlyAsnIleLeuGlyHis 139
 QY 415 AAGACCATGGCTGGAGGCTCCACCGAGCGCTGTAC-----CCCCGC 459
 Db 140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
 QY 460 GACGCGTGTGAAGGGGAGATCACCAAGCCCTGAAGCTGAAGGACGCGGCCACTAC 519
 Db 159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
 QY 520 CTGTGGAGTTC-----AAGTCCATCTACATGCGCAAGAGCCGCTGACGCGCGC 573
 Db 178 LeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp 197
 QY 574 TACTACTACGTGAGCTACGAGCGCAGCGAGCGCGCGCCACAC 666
 Db 198 AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
 QY 625 ACCATCTGGAGCTACGAGCGCAGCGAGCGCGCGCCACAC 666
 Db 218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231
 RESULT 12
 US-08-819-612-22
 ; Sequence 22, Application US/08819612D
 ; Patent No. 6172188
 ; GENERAL INFORMATION:
 ; APPLICANT: Thastrup, et al.
 ; TITLE OF INVENTION: No. 6172188el Fluorescent Proteins
 ; FILE REFERENCE: No. 6172188el Fluorescent Proteins
 ; CURRENT APPLICATION NUMBER: US/08/819.612D
 ; CURRENT FILING DATE: 1997-03-17
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 22
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 US-08-819-612-22

Alignment Scores:
 Pred. No.: 3,27e-12 Length: 238
 Score: 226.50 Matches: 60
 Percent Similarity: 47.44% Conservatives: 51
 Best Local Similarity: 25.64% Mismatches: 108
 Query Match: 18.31% Indels: 15
 DB: 3 Gaps: 7
 US-10-081-864-14 (1-678) x US-08-819-612-22 (1-238)

QY 1 ATGGCTCTCCGAGAACGTCATCCAGGTTTCATGGCTTCAAGTGCGCATGGAGGC 60
 Db 1 MetSerLysGlyGluGluPheThrGlyValValProIleLeuValGluLeuAspGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGCGCCCTACGAGGC 120
 Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGlyGluGlyAspAlaThrTyrGly 40
 QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGGCGGCCCTCGCTTCGCTGGGACATC 180
 Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
 QY 181 CTGTCCTCCAGTTCACGCTCCAGGCTCCAGGCGGCTTCAAGTGGAGCGCGTGAAC 240
 Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
 QY 295 TTCGAGACGGCGCGTGGCGACCGTGAACCGAGCTCTCCCTCGAGGCGGCTCTTC 354
 Db 100 TyrLysAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119

QY 355 ATCTACAAGTGAAGTTTCATCGGCTGAACCTCCCTCCGAGGCGCGCGTGCAGAG 414
 Db 120 ValAsnArgIleGluLeuLysGlyLeuAspPheLysGluAspGlyAsnIleLeuGlyHis 139
 QY 415 AAGACCATGGCTGGAGGCTCCACCGAGCGCTGTAC-----CCCCGC 459
 Db 140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
 QY 460 GACGCGTGTGAAGGGGAGATCACCAAGCCCTGAAGCTGAAGGACGCGGCCACTAC 519
 Db 159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
 QY 520 CTGTGGAGTTC-----AAGTCCATCTACATGCGCAAGAGCCGCTGACGCGCGC 573
 Db 178 LeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp 197
 QY 574 TACTACTACGTGAGCTACGAGCTGACATCAC-----TCCACACGAGGACTAC 624
 Db 198 AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
 QY 625 ACCATCTGGAGCTACGAGCGCAGCGAGCGCGCGCCACAC 666
 Db 218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 13
 US-09-316-919-3
 ; Sequence 3, Application US/09316919
 ; Patent No. 6469154
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Baird, Geoffrey
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
 ; FILE REFERENCE: 07257/073001
 ; CURRENT APPLICATION NUMBER: US/09/316.919
 ; CURRENT FILING DATE: 1999-05-21
 ; NUMBER OF SEQ ID NOS: 63
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 US-09-316-919-3

Alignment Scores:
 Pred. No.: 3,27e-12 Length: 238
 Score: 226.50 Matches: 60
 Percent Similarity: 47.44% Conservatives: 51
 Best Local Similarity: 25.64% Mismatches: 108
 Query Match: 18.31% Indels: 15
 DB: 4 Gaps: 7
 US-10-081-864-14 (1-678) x US-09-316-919-3 (1-238)

QY 1 ATGGCTCTCCGAGAACGTCATCCAGGTTTCATGGCTTCAAGTGCGCATGGAGGC 60
 Db 1 MetSerLysGlyGluGluPheThrGlyValValProIleLeuValGluLeuAspGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGCGCCCTACGAGGC 120
 Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGlyGluGlyAspAlaThrTyrGly 40
 QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGGCGGCCCTCGCTTCGCTGGGACATC 180
 Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
 QY 181 CTGTCCTCCAGTTCACGCTCCAGGCTCCAGGCTCCAGGCGGCTTCAAGTGGAGCGCGTGAAC 240
 Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
 QY 241 -----GACTACAGAAGCTCTCTCCCGAGGCGCTTCAAGTGGAGCGCGTGAAC 294
 Db 80 GlnHisAspPheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99

```
QY 295 TTGAGGACGGGGGTGGCGACCGTGCAGCCAGGACTCCTCCCTGCAGGACGGCTGCTTC 354
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 ATCTACAGGTGAAGTTCATCGCGGTGAATTCCTCCGACGGCCCGGTGATGCAGAG 414
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 415 AAGACCATGGGTGGGAGCGCTCCACCGAGCGCTGTAC-----CCCGCG 459
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 460 GACGGGTCTGAGGGCGAGATCCACAGGCGCTGAGCTGAAGACGGCGGCGCCACTAC 519
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 520 CTGGTGGAGTTC-----AAGTCCATCTACATGGCCAGAGCGCTGCAGCTGCGCGCG 573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 LeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 574 TACTACTAGTGACACCAAGCTGGACATCAC-----TCCACACAGAGACTAC 624
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 AsnHisTyrSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 625 ACCATCTGAGCAGTACGAGCGCACCGAGGGCGGCCACCAC 666
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231
```

RESULT 14

US-09-346-946-32
; Sequence 32, Application US/09346946
; Patent No. 6566083

GENERAL INFORMATION:
; APPLICANT: Thastrup, Ole
; APPLICANT: Tullin, Soren
; APPLICANT: Poulsen, Lars
; APPLICANT: Bjorn, Sara

; TITLE OF INVENTION: A Method of Detecting Biologically
; FILE REFERENCE: 4301.204-US
; CURRENT APPLICATION NUMBER: US/09/346,946
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US/08/818,604
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 0110/95
; PRIOR FILING DATE: 1995-01-31
; PRIOR APPLICATION NUMBER: 0982/95
; PRIOR FILING DATE: 1995-09-07
; PRIOR APPLICATION NUMBER: PCT/DK96/00052
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequora victoria
US-09-346-946-32

Alignment Scores:

Pred. No.:	3,27e-12	Length:	238
Score:	226.50	Matches:	60
Percent Similarity:	47.44%	Conservative:	51
Best Local Similarity:	25.64%	Mismatches:	108
Query Match:	18.31%	Indels:	15
DB:	4	Gaps:	7

US-10-081-864-14 (1-678) x US-09-346-946-32 (1-238)

```
QY 1 ATGGCTCTCCGAGACGTCATCACCGAGTTCAGCTTCAGGTGCGCATGGAGGCG 60
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 61 ACCGTGAACCGCCACGAGTTCGAGATCGAGGGCGAGGGCCCGCCCTACGAGGCG 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGluGlyAspAlaThrTyrGly 40
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 CACAACACCGTGAAGTTGAAGTGCACCAAGGGGGGGCCCTGCCTTCGCTGGACATC 180
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 CTGTCCCGCCAGTTCAGTTCAGGTCCAGGTGATCAGTGAAGACACCCCGCGGACATCCC 240
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 -----GACTACAGAAGAGTCTCTCTCCCGAGGGCTTCAAGTGGGAGCGCTGATCAAC 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 GlnHisAspPhePheLysSerAlaMetProIleGlyTyrValGlnGluAspThrIlePhe 99
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 TTCGAGGACGGCGCGTGGCGCACCGTGCACCCAGAGACTCTCCCTGCAGGACGGCTGCTTC 354
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 ATCTACAGGTGAAGTTCATCGCGGTGAATTCCTCCGACGGCCCGGTGATGCAGAG 414
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 415 AAGACCATGGGTGGGAGCGCTCCACCGAGCGCTGTAC-----CCCGCG 459
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 460 GACGGGTCTGAGGGCGAGATCCACAGGCGCTGAGCTGAAGACGGCGGCGCCACTAC 519
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 520 CTGGTGGAGTTC-----AAGTCCATCTACATGGCCAGAGCGCTGCAGCTGCGCGCG 573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 LeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 574 TACTACTAGTGACACCAAGCTGGACATCAC-----TCCACACAGAGACTAC 624
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 AsnHisTyrSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 625 ACCATCTGAGCAGTACGAGCGCACCGAGGGCGGCCACCAC 666
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231
```

RESULT 15

US-09-023-946B-22
; Sequence 22, Application US/09023946B
; Patent No. 6670449

GENERAL INFORMATION:
; APPLICANT: GERO MISENBOCK, ET AL.

; TITLE OF INVENTION: HYBRID MOLECULES AND
; THEIR USE FOR OPTICALLY DETECTING CHANGES IN
; CELLULAR MICROENVIRONMENTS

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,946B

FILING DATE: 13-Feb-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/036,805

FILING DATE: 14 FEBRUARY 1997

APPLICATION NUMBER: 60/038,179


```
; FILING DATE: 13 FEBRUARY 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 2955-4004US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-023-946B-22
```

```
Alignment Scores:
Pred. No.: 3.27e-12 Length: 238
Score: 226.50 Matches: 60
Percent Similarity: 46.78% Conservative: 49
Best Local Similarity: 25.75% Mismatches: 111
Query Match: 18.31% Indels: 13
DB: 4 Gaps: 6
```

US-10-081-864-14 (1-678) x US-09-023-946B-22 (1-238)

```
QY 1 ATGCGCTCTCCGAGACGTCATCCAGCATTCATCGCTTCAAGTGGCGATGGAGGC 60
Db 1 MetSerLysGlyGluGluLeuPheThrGlyValValProLeuLeuGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGGCCCTACGAGGGC 120
Db 21 AspValAsnGlyHisLysPheSerValSerGlyGluGluGlyAspAlaThrTyrGly 40
QY 121 CACAACACCGTGAAGTTCAGAGTGAAGCCAGGGCGGCCCTTGCCTTGGCTGGGACATC 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProTyrProThr 59
QY 181 CTGTCGCCCGCAGTTCAGTCCAGGTCCAGGTGAGTGAAGCAGCCCGCGACATCCCC 240
Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
QY 241 -----GACTACAAAGAGTGTCTTCCCGAGGGTTCAGTGGGAGCGCGTGAAC 294
Db 80 ArgHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99
QY 295 TTCGAGGACGGCGGTGGCGACCGTGCACCGAGCTCTCCCTGCGAGGCGGCTCTTC 354
Db 100 PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAGGTGAAGTTCATCGCGGTCACTTCCCTCCGAGCGGCCCGCGTGAAG 414
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
QY 415 AAGACCATGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGCGACGCGCTGTGAAG 474
Db 140 Lys---LeuGluTyrAsnTyrAsnAspHisGlnValTyrIleMetAlaAspLysGlnLys 158
QY 475 GCGGAGATCCACAGGCGCTGAAGCTGAAG-----GACGCGCGCCACTACCTG 522
Db 159 AsnGlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlyValGlnLeu 178
QY 523 GTGAGTTC-----AAGTCCATCTACATGGCCAGAGCGCGTGCAGCTGCCCGCTAC 576
Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198
QY 577 TACTACGTGGACACCAAGCTGGACATCAC-----TCCCAACAGGAGGACTACCC 627
Db 199 HisTyrLeuPheThrThrSerThrLeuSerLysAspProAsnGlnLysArgAspHisMet 218
QY 628 ATCGTGAGCAGTACGAGCAGCGAGGCGCGCCACCCAC 666
```

```
Db 219 ValLeuLeuGluPheValThrAlaAlaGlyIleThrHis 231
Search completed: July 29, 2004, 14:38:33
Job time : 17.9516 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:36:21 ; Search time 43.0787 Seconds
(without alignments)
9873.863 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggcctctctcgagacgt.....gccaccacctgtctctgtaa 678

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 2582470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10081864/runat_29072004_150508_25599/app_query.fasta_1.1742
-DB=Published Applications AA -QWTF=fastan -SUFFIX=xapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXI=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USRP=US10081864 @CGN 1.1 21 @runat_29072004_150508_25599
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCX=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	98.1	1214	US-10-315-920-6	Sequence 6, Appl
2	97.3	1204	US-10-442-148A-7	Sequence 7, Appl
3	97.3	1204	US-10-442-148A-8	Sequence 8, Appl
4	96.9	1199	US-10-315-920-4	Sequence 4, Appl
5	96.3	1191	US-09-999-745-67	Sequence 67, Appl
6	96.3	1191	US-09-866-538-12	Sequence 12, Appl
7	96.3	1191	US-09-794-308-12	Sequence 12, Appl
8	96.3	1191	US-09-865-291-12	Sequence 12, Appl
9	96.3	1191	US-10-132-067-4	Sequence 4, Appl
10	96.3	1191	US-10-006-922-12	Sequence 8, Appl
11	96.3	1191	US-10-081-864-8	Sequence 1, Appl
12	96.3	1191	US-10-121-258-1	Sequence 2, Appl
13	96.3	1191	US-10-315-920-2	Sequence 56, Appl
14	96.3	1191	US-10-370-570-56	Sequence 32, Appl
15	96.3	1191	US-10-406-618-32	Sequence 13, Appl
16	96.3	1191	US-10-433-640-13	Sequence 1, Appl
17	96.3	1191	US-10-343-977-1	Sequence 2, Appl
18	96.3	1191	US-10-343-977-2	Sequence 3, Appl
19	96.3	1191	US-10-343-977-3	Sequence 2, Appl
20	96.0	1187	US-10-152-296-2	Sequence 2, Appl
21	96.0	1187	US-10-739-656-2	Sequence 2, Appl
22	96.0	1187	US-10-214-932-52	Sequence 52, Appl
23	96.0	1187	US-10-214-932-76	Sequence 76, Appl
24	95.7	1184	US-10-121-258-20	Sequence 20, Appl
25	95.5	1181	US-10-314-936-2	Sequence 2, Appl
26	95.5	1181	US-10-314-936-4	Sequence 4, Appl
27	94.8	1172.5	US-10-724-178-12	Sequence 12, Appl
28	93.5	1156	US-10-121-258-4	Sequence 4, Appl
29	92.9	1149	US-10-121-258-24	Sequence 24, Appl
30	91.4	1130	US-10-423-688A-41	Sequence 41, Appl
31	90.9	1124	US-10-121-258-6	Sequence 6, Appl
32	89.2	1104	US-10-006-922-44	Sequence 44, Appl
33	89.2	1104	US-10-081-864-12	Sequence 12, Appl
34	86.5	1070.5	US-10-006-922-18	Sequence 18, Appl
35	86.5	1070.5	US-10-161-403-40	Sequence 40, Appl
36	82.7	1023	US-10-006-922-46	Sequence 46, Appl
37	82.5	1021	US-10-121-258-8	Sequence 8, Appl
38	82.5	1021	US-10-724-178-16	Sequence 16, Appl
39	84.6	846	US-10-724-178-1041	Sequence 1041, Ap
40	83.2	832	US-10-724-178-1033	Sequence 1033, Ap
41	79.3	793	US-10-724-178-1037	Sequence 1037, Ap
42	76.2	762	US-10-006-922-10	Sequence 10, Appl
43	76.2	762	US-10-724-178-1033	Sequence 1033, Ap
44	58.7	726.5	US-10-133-973-5	Sequence 5, Appl
45	58.7	726.5	US-10-370-570-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1
US-10-315-920-6
; Sequence, Application US/10315920
; Publication No. US20030175609A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.

US-10-315-920-6

Alignment Scores:

Pred. No.: 7,38e-84 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.14% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-14 (1-678) x US-10-315-920-6 (1-225)

```
QY 1 ATGGCTCTCCGAGACGTCATCAGGCTTCATGCGCTTCAAGTGCATGAGGC 60
DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGCGGAGGCGGCCGCCCTTACGAGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyA-gProTyrGluGly 40
QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGCGGCGGCCCTTCCCTTCGCTCGGACATC 180
DB 41 HisAsnThrValLysLeuSerPheProGluGlyPheLysValArgMetAsnPheGlu 60
QY 181 CTGTCCCCCAGTTCACGCTCCAGGTGACGAGTGTACGTGAAGCACGCCCGCGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGTGTCTCTCCCGAGGGTTCAGTGGAGCGCGTGAATCTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysValTyrValLysHisProAlaAspIlePro 100
QY 301 GACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 101 AspGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 120
QY 361 AAGTGAAGTTCATCGCGGTGAAGTTCCTCCCGAGCGGCGGCGGCGGCGGCGGCGGCGG 420
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCAGCGGCGGCGG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTCTCTG 675
DB 221 HisHisLeuPheLeu 225
```

RESULT 2

US-10-442-148A-7

; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO
; APPLICANT: HIROTA, KAYUNORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583.0103-00000
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950

; PRIOR FILING DATE: 2002-05-23

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7

; LENGTH: 225

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

Alignment Scores:

Pred. No.: 4,24e-83 Length: 225
Score: 1204.00 Matches: 223
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 2
Query Match: 97.33% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-14 (1-678) x US-10-442-148A-7 (1-225)

```
QY 1 ATGGCTCTCCGAGACGTCATCAGGCTTCATGCGCTTCAAGTGCATGAGGC 60
DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGCGGAGGCGGCCGCCCTTACGAGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyA-gProTyrGluGly 40
QY 121 CACACACCGTGAAGTTCAGGTGACCAAGGCGGCGGCCCTTCCCTTCGCTCGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCCCAGTTCACGCTCCAGGTGACGAGTGTACGTGAAGCACGCCCGCGGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATCTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysValTyrGluArgValMetAsnPheGlu 100
QY 301 GACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGCGGTGAAGTTCCTCCCGAGCGGCGGCGGCGGCGGCGGCGGCGG 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGCGGCGGCGGCGGCGGCGGCGGCG 480
DB 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
DB 161 ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCAGCGGCGGCGG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTCTCTG 675
DB 221 HisHisLeuPheLeu 225
```

RESULT 3

US-10-442-148A-8

; Sequence 8, Application US/10442148A
; Publication No. US20040014242A1

GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHITO
; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583.0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-8

Alignment Scores:
Pred. No.: 4,28e-83 Length: 239
Score: 1204.00 Matches: 223
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 2
Query Match: 97.33% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-14 (1-678) x US-10-442-148A-8 (1-239)

```
QY 1 ATGGCTCTCTCCGAGACGTCATCACCAGTTCATCGGCTTCAAGTGGCGATGAGGGC 60
DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAAGCCGACGAGTTCAGATCGAGGGCGAGGGCGAGGGCGGCGCCCTACAGGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAACAGGGCGGCGCCCTGCGCTTCCGCTGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAAGTGTAGTGAAGCACCGCCCGACATCCCG 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGGGCGTGGAGCGTCCACCGAGCGCTGTACCCCGCGGCGCCCGTGTATGAGAGAGAC 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGGACGCGCTGTGAGAGCGGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGCCCTGAAGCTGAAGGAGCGGCGGCGCTACTGTGTGGAGTTCAAGTCCATC 540
DB 161 ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAGCCCGTGCAGCTCCCGGCTACTACTAGTGTGACACCAAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspAlaLysLeuAsp 200
QY 601 ATCACTCCACACGAGGACTACACATCGTGGAGCAGTACGAGCGCACCGGCGGCGC 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
```

```
QY 661 CACCACCTGTTCCTG 675
DB 221 HisHisLeuPheLeu 225
RESULT 4
US-10-315-920-4
; Sequence 4, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-4
```

Alignment Scores:
Pred. No.: 1,02e-82 Length: 225
Score: 1199.00 Matches: 222
Percent Similarity: 99.11% Conservative: 1
Best Local Similarity: 98.67% Mismatches: 2
Query Match: 96.93% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-14 (1-678) x US-10-315-920-4 (1-225)

```
QY 1 ATGGCTCTCTCCGAGACGTCATCACCAGTTCATCGGCTTCAAGTGGCGATGAGGGC 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAAGCCGACGAGTTCAGATCGAGGGCGAGGGCGAGGGCGGCGCCCTACAGGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAACAGGGCGGCGCCCTGCGCTTCCGCTGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAAGTGTACGTGAAGCACCGCCCGACATCCCG 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGGGCGTGGAGCGTCCACCGAGCGCTGTACCCCGCGGCGCCCGTGTATGAGAGAC 360
DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGGCGTGAACCTTCCCTCCGAGCGGCGGCGGTGATGACGAAGAGAC 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGGCGGCGCTGCTGAAGGGCGGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
```

QY 481 ATCCACAGGCCCTGAAGCTGAAGAGCGGGCCACTACTCTGTGGAGTTCAAGTCCATC 540
 DB 161 ILeHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGCCAGAGCCCGTGCAGCTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
 QY 601 ATCACTCCACACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlyTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTTCTCTG 675
 DB 221 HisHisLeuPheLeu 225
 RESULT 5
 US-09-999-745-67
 ; Sequence 67, Application US/09999745
 ; Patent No. US20020157120A1
 ; GENERAL INFORMATION:
 ; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Baird, Geoffrey
 ; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
 ; FILE REFERENCE: REGEX1470-1
 ; CURRENT APPLICATION NUMBER: US/09/999,745
 ; CURRENT FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: 09/316,920
 ; PRIOR FILING DATE: 1999-05-21
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 67
 ; LENGTH: 225
 ; TYPE: PRT
 ; ORGANISM: Discosoma sp.
 US-09-999-745-67

Alignment Scores:
 Pred. No.: 4.12e-82 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.67% Conservative: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 9 Gaps: 0

US-10-081-864-14 (1-678) x US-09-999-745-67 (1-225)

QY 1 ATGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGCATGGAGGC 60
 DB 1 MetArgSerSerLysAsnValIleLysGluPheValArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAACACCGTGAAGTTGAAGTGCACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCCGCCAGTTCCTTCCCGGAGGCTTCAAGTGGAGCGGCGGCGGCGGCGGCGGCGG 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAGAAGCTGTCTTCCCGGAGGCTTCAAGTGGAGCGGCGGCGGCGGCGGCGGCGG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100
 QY 301 GACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 DB 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGTGAAGTTTCATCGGCGGTGAACCTTCCCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 420

DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGCTCGGAGGCGCTCCACCGAGCGCTGTACCCCGCGGCGGCGGCGGCGGCGGCGGCGG 480
 DB 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCACAGGCCCTGAAGCTGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 DB 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGCCAGAGCCCGTGCAGCTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATCACTCCCAACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGlnGlyTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTTCTCTG 675
 DB 221 HisHisLeuPheLeu 225
 RESULT 6
 US-09-866-538-12
 ; Sequence 12, Application US/09866538
 ; Publication No. US20030032088A1
 ; GENERAL INFORMATION:
 ; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
 ; APPLICANT: Tsien, Roger
 ; APPLICANT: Campbell, Robert
 ; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
 ; FILE REFERENCE: REGEX1530-2
 ; CURRENT APPLICATION NUMBER: US/09/866,538
 ; CURRENT FILING DATE: 2001-05-24
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 225
 ; TYPE: PRT
 ; ORGANISM: Discosoma sp.
 US-09-866-538-12

Alignment Scores:
 Pred. No.: 4.12e-82 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.67% Conservative: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 10 Gaps: 0

US-10-081-864-14 (1-678) x US-09-866-538-12 (1-225)

QY 1 ATGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGCATGGAGGC 60
 DB 1 MetArgSerSerLysAsnValIleLysGluPheValArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAACACCGTGAAGTTGAAGTGCACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCCGCCAGTTCCTTCCCGGAGGCTTCAAGTGGAGCGGCGGCGGCGGCGGCGGCGG 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAGAAGCTGTCTTCCCGGAGGCTTCAAGTGGAGCGGCGGCGGCGGCGGCGGCGG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100

Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrVallYshHisProAlaAspIlePro 80
Qy 241 GACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCTGATGAATTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGCGGCTGCGACCGGTGACCCAGGACTCTCTCCCTGACGACGCGCTGCTCATCTAC 360
Db 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTGAAGTTCATCGCGGAACTTCCCTCCGACGCGCGCTGATGCGAAGAGAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGCGCTGCTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATCCACAAAGCGCTTGAAGTGAAGCGCGGCGCTTACCTGCTGAGTTCATCTCATC 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGCCCAAGAACCGCTGCGAGCTGCGCGCTACTACTACTGCTGAGTTCATCTCATC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATCACCCTCCACACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCACCTGTCTCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 8

US-09-865-291-12
; Sequence 12, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-865-291-12

Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 10 Gaps: 0

US-10-081-864-14 (1-678) x US-09-865-291-12 (1-225)

Qy 1 ATGGCTCTCCGAGAGCTCATCACCGAGTTCATGCGCTTCAAGTTCGATGAGGCG 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACCGTGAAGCGCCCGAGTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGG 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyGluGlyGluGly 40

Qy 301 GACGGCGGCTGCGACCGGTGACCCAGGACTCTCTCCCTGACGACGCGCTGCTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTGAAGTTCATCGCGGAACTTCCCTCCGACGCGCGCTGATGCGAAGAGAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGCGCTGCTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATCCACAAAGCGCTTGAAGTGAAGCGCGGCGCTTACCTGCTGAGTTCATCTCATC 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGCCCAAGAACCGCTGCGAGCTGCGCGCTACTACTACTGCTGAGTTCATCTCATC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATCACCCTCCACACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCACCTGTCTCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 7

US-09-794-308-12
; Sequence 12, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-794-308-12

Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 10 Gaps: 0

US-10-081-864-14 (1-678) x US-09-794-308-12 (1-225)

Qy 1 ATGGCTCTCCGAGAGCTCATCACCGAGTTCATGCGCTTCAAGTTCGATGAGGCG 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACCGTGAAGCGCCCGAGTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyGluGlyGluGly 40
Qy 121 CACACACCGTGAAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 CTGTCCCGGCGGCTCCAGTTCAGGCTCCAGGCTGACGAGTTCAGGCTGACGAGTTCAGGCTG 240

QY 121 CACACACCGTGAAGTTGAAGTACCAAGCGCGCCCTGCGCTTCGCTGGACATC 180
Db 41 HisasnThrValLysLeuLysValThrLysGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCTCCCGGAGTTCAGTACGGTCCCAAGGTGTACGTGAAGACACCCCGCGCATCCCC 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGAGCGCGGTGATGAAGTTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGCGCGCGTGGCGACCGTACCGAGACTCTCCCTCCGAGGAGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGCGTGAAGTTCCTCCGAGCGCGCCCGTGTGATGCAAGAAAGACC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAAGCGCTGAAGTGAAGGACGGCGGCACTACCTGTGTGAGTTCAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAGAGCGCGTGCAGCTCCCGGCTACTACTACCTGACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACTCCACACAGGAGTACACCATCGTGAGCAGTACAGCGCGCACCGAGGGCGCG 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 9
US-10-132-067-4
; Sequence 4, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
; TITLE OF INVENTION: Fluorescence
; FILE REFERENCE: 021362-000600US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; OTHER INFORMATION: red fluorescent protein (dsRED)
US-10-132-067-4

Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 12 Gaps: 0

US-10-081-864-14 (1-678) x US-10-132-067-4 (1-225)

QY 1 ATGGCCTCTCTCGAAGACGTGATCACCAGAGTTCATGCGCTTCAAGTTCGCGATGAGGGC 50
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGCGGAGGCGGCGCCCTTACGAGGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGlyGly 40
QY 121 CACACACCGTGAAGTTCAGGTGACCAAGGGCGGCGCCCTTCCCTGCGCTGGAGATC 180
Db 41 HisasnThrValLysLeuLysValThrLysGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCTCCCGGAGTTCAGTACGGTCCCAAGGTGTACGTGAAGACACCCCGCGCATCCCC 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAGTTCCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGTGATGAAGTTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGCGCGCGTGGCGACCGTACCGAGACTCTCCCTCCGAGGAGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGCGTGAAGTTCCTCCGAGCGCGCCCGTGTGATGCAAGAAAGACC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAAGCGCTGAAGTGAAGGACGGCGGCACTACCTGTGTGAGTTCAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAAGACCGCGTGCAGCTCCCGGCTACTACTACCTGGAACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACTCCACACAGGAGTACACCATCGTGAGCAGTACAGCGCGCACCGAGGGCGCG 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 10
US-10-006-922-12
; Sequence 12, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Pradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556

Appl.


```

; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Bradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR FILING DATE: 2001-12-04
; PRIOR FILING DATE: 2001-12-04
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; US-10-081-864-8

Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-14 (1-678) x US-10-081-864-8 (1-225)

QY 1 ATGGCCCTCCCGAGACGTCATCCAGGTTTCATCGGCTTCAAGTGGCGATGAGGCG 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCAGATCGAGCGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTCAGATCGAGCGGGGGGGGGGGGGGGGGGGGGGGGG 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCTCCCGAGTTCAGATCGAGCGGGGGGGGGGGGGGGGGGGGGGGGG 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGTTCCTTCCCGAGGCTTCAAGTGGCGGGGGGGGGGGGGGGGG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 360
DB 101 AspGlyGlyValValThrValThrGluAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGGCGTGAATTCCTCCCGAGGCGGGGGGGGGGGGGGGGG 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCGTGGAGGCGCTCCACCGAGCGGTGTACCCCGCGAGCGGGGGGGGGGG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGGCGCTGAAGTGAAGGACGGCGGGGGGGGGGGGGGGGGGGGGGG 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAAGAGCGCGGTGCGGCTGCTACTACTACTACTACTACTACTACT 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

; RESULT 11
; US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication No. US20030022287A1

```


Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGGACCGTGCACCCAGGACTCTCCCTCCGAGGAGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCAATCGGCGTGAACCTCCCTCCGAGGAGCGGCTGCTTCATCTAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGAGCGGCTGCTGAAGGCGGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGCCCTGAAGCTGAAGGACGCGCGCCCACTACCTGGTGGAGTTCAAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAAGCCGCTGAGCTGCCCGCTACTACTAGCTGGACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACCCTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGAGCGCACCAAGCTGGAC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 14
US-10-370-570-56
; Sequence 56, Application US/10370570
; Publication No. US2003021971A1
; GENERAL INFORMATION:
; APPLICANT: DAHL, Soren Weis et al.
; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS
; FILE REFERENCE: 3759-0130P
; CURRENT APPLICATION NUMBER: US/10/370,570
; CURRENT FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 56
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Diatoms sp.
US-10-370-570-56

Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-14 (1-678) x US-10-370-570-56 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTTCAGGCTTCAAGTGGCGATGAGGAGGC 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTTCAGATCGAGGCGGAGGCGGCGGCGGCTTCAAGGAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAGGTGACCAAGGCGGCGGCGGCTTCCGCTGGGACATC 180
Db 41 HisAsnThrValLysLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGGCGGAGTTCAGTACGCTCAAGGTGTAGTGAAGCACCCGCGGACATCCCG 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

QY 241 GACTACAGAAGCTGTCTTCCCGAGGCGCTTCAAGTGGAGCGGCTGATGAAGTCTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCTGGCGACCGTGCACCCAGGACTCTCCCTCCGAGGAGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCAATCGGCGTGAACCTCCCTCCGAGGAGCGGCTGCTTCATCTAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGAGCGGCTGCTGAAGGCGGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGCCCTGAAGCTGAAGGACGCGCGCCCACTACCTGGTGGAGTTCAAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAAGCCGCTGAGCTGCCCGCTACTACTAGCTGGACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACCCTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGAGCGCACCAAGCTGGAC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 15
US-10-406-618-32
; Sequence 32, Application US/10406618
; Publication No. US2003021981A1
; GENERAL INFORMATION:
; APPLICANT: Wan, David Chi-Cheong
; APPLICANT: Ip, Denis Tsz-Ming
; APPLICANT: The Chinese University of Hong Kong
; TITLE OF INVENTION: NO. US2003021981A1el Fluorescent Proteins
; FILE REFERENCE: 016285-34-1
; CURRENT APPLICATION NUMBER: US/10/406,618
; CURRENT FILING DATE: 2003-04-02
; PRIOR APPLICATION NUMBER: US 60/387,968
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: US 60/370,598
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 32
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Spontaneously
; OTHER INFORMATION: Fluorescent protein DsRed.
US-10-406-618-32

Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-14 (1-678) x US-10-406-618-32 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTTCAGGCTTCAAGTGGCGATGAGGAGGC 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

```
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTACGAGGGC 120
Db |||
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyArgProTyrGluGly 40
Db |||
QY 121 CACAACACCGTGAAGTTGAAGTACCAAGGGGGCCCTGCCTTCGCTCGGACATC 180
Db |||
QY 41 HisAsnThrValLysLeuLysValThrLysGlyProLeuPheAlaTrpAspIle 60
QY 181 CTGTCCCCCAGTTCCTAGTACGGCTCCAAGGTGACGAGACACCCCGCGACATCCCC 240
Db |||
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATTCGAG 300
Db |||
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGCGGTGGCGACCGCTGACCCAGGACTCTCTCCCTGCAGGACGGCTCTTCATCTAC 360
Db |||
QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGCGCTGAAGTCCCTCCGACGGCCCGGTGATGACAGAGAGACC 420
Db |||
QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGGTGCTGAAGGGCGAG 480
Db |||
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGCCCTGAAGCTGAAGACGGCGGCCACTACCTGTGGAGTTCAGTCCATC 540
Db |||
QY 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAAGACCGCTGCAGCTGCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
Db |||
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACCTCCACAAACGAGGACTACCATCGTGGAGCAGTACGAGCGCACCGAGGCCCGC 660
Db |||
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTCTCTG 675
Db |||
QY 221 HisHisLeuPheLeu 225
```

Search completed: July 29, 2004, 14:54:49
Job time : 46.0787 secs

OM nucleic - protein search, using frame_plus_n2p model
Run on: July 29, 2004, 14:28:05 ; Search time 14.6859 Seconds
(without alignment)
8881.684 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggcctctccgagaagct.....gccaccacctgttctgttaa 678

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2.1/USPTO.spool_p/US10081864/runat 29072004 150506 25492/app query.fasta_1.1742
-DB=PIR.78 -QPM=fastan -SUFFIX=pr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 @CIGN 1.1 53 @runat 29072004 150506 25492 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DVT TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR.78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225.5	18.2	238	JQ1514	green-fluorescent
2	205.5	16.6	580	T43481	probable mucin DKF
3	203	16.4	3020	A43932	mucin 2 precursor,
4	186	15.0	1106	JQ4003	hypothetical 119.5
5	182	14.6	1791	T02345	hypothetical prote
6	175.5	14.2	279	S53363	mucin SAC (clone J
7	175	14.0	660	1 QQB3	BHLF1 protein - hu
8	174	14.1	317	S55316	mucin (clone PGM-2
9	171	13.8	377	A48018	mucin 7 precursor,
10	170.5	13.8	528	T47141	gastric mucin (clo
11	170.5	13.8	1367	1 S48478	glucan 1,4-alpha-g
12	168.5	13.6	335	T05722	licheninase (EC 3.
13	165.5	13.4	1161	2 S57180	probable membrane
14	164	13.3	383	2 T46707	proteophosphoglyca

15	163	13.2	507	2 T44768	antifreeze glycope
16	163	13.0	749	2 A45294	Balbani ring 2.1
17	162.5	13.1	283	2 S13383	hydroxyproline-ric
18	160	12.9	839	2 T04859	extensin homolog F
19	159.5	12.9	267	2 S08314	cell wall glycopro
20	159.5	12.9	529	2 T45134	hypothetical prote
21	159	12.9	464	2 S22697	extensin - Volvox
22	157	12.7	350	2 S22456	hydroxyproline-ric
23	157	12.7	494	2 G84348	hypothetical prote
24	156	12.6	525	2 JN0059	hypothetical 57.4K
25	154.5	12.4	221	2 C34768	ORF2 protein - Orf
26	154.5	12.5	866	2 T45462	membrane glycoprot
27	153	12.2	414	2 C86301	arginine/serine-ri
28	153	12.4	770	2 T22808	hypothetical prote
29	152.5	12.3	164	2 T53641	mucin 5AC - human
30	152	12.3	264	2 S29893	salivary glue prot
31	152	12.3	303	2 S28364	hydroxyproline-ric
32	152	12.3	867	2 T45463	membrane glycoprot
33	151.5	12.2	797	1 VGBEX1	glycoprotein X pre
34	150.5	12.2	222	2 T96711	hypothetical prote
35	150.5	12.2	477	2 S53362	mucin 5AC (clone J
36	150.5	12.0	1520	2 T44231	hypothetical prote
37	150.5	12.2	3570	2 T45025	mucin MUC5B, trach
38	150	12.1	328	2 TQ0985	hydroxyproline-ric
39	149.5	12.0	528	2 B75310	conserved hypotnet
40	149.5	12.0	1513	2 A54895	mucin 2, intestina
41	148.5	12.0	319	2 F75420	hypothetical prote
42	148	11.8	322	2 A53715	apomucin precursor
43	148	12.0	214	2 T10737	extensin-like cell
44	147.5	11.9	214	2 T10737	chitinase (EC 3.2.
45	147.5	11.9	660	2 JN0067	

ALIGNMENTS

RESULT 1
JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C>Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:11347277
A:Accession: JQ1514
A:Molecule type: DNA
A:Residues: 1-107,'S',109-238 <PRA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64,74-122,132-151,154-183,185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>
A:Cross-references: GB:I29345; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'R',
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1

QY 263 CCGAGGCTTCAAGTGGAGCGGTGATGAACTTCGAGACGCGCGTGGCGACCGTGA 322
Db 173 ThrAlaSerLeuThrArgThr-----ProSerArgAlaSerLeuThrArgTrpPro 189
QY 323 CCCAGGACTCCTCCCTGCGAGGACG-----OCTGCT 352
Db 190 ProArgAlaSerProThrArgThrProProArgGluSerProArgMetSerHisArgAla 209
QY 353 TCATCTACAGGTGAAGTTCATCGGCGTGAATCTCCCTCCGACGCGCCCGTATGACAGA 412
Db 210 SerProThrArg-----ThrProProArgAlaSerProThrArg 222
QY 413 AGAAGACCATGGGCTGGG-----AGGCCTCCACCG-----AGC 445
Db 223 ArgProProArgAlaSerProThrArgThrProProArgGluSerLeuAthrSerHis 242
QY 446 GCGTGTACCCCGGACGCGGTGTAAGGGGAGATCCACAGG----- 490
Db 243 ArgAlaSerProThrArgMetProProArgAlaSerProThrArgArgProProArgAla 262
QY 491 -----CCCTGAGCTGAAGACGCGCGCCACCTACCTGCTGGAGTTCA 532
Db 263 SerProThrGlySerProPro-----ArgAlaSerProMetThrProProArgAla 279
QY 533 AGTCCATCTACATGGCCAAAGA---AGCCGTGACGCTGCCG-----GCTACTACTACG 583
Db 280 SerProArgThrProProArgAlaSerProThrThrProSerArgAlaSerLeuThr 299
QY 584 TGGACACGAGTGGACATCCTCCACA----- 613
Db 300 ArgThrProSerTrpAlaSerProThrThrProSerArgAlaSerLeuMetLysMet 319
QY 614 -----ACGAGGACTACACCATCGTGGAGCAGTACGACGCGCACCG--- 652
Db 320 GluSerThrValSerIleThrArgThrProProArgAlaSerProThrGlyThrProSer 339
QY 653 AGGCGCGCACCAACC 667
Db 340 ArgAlaSerProThr 344

RESULT 3
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin 5MUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999
C;Accession: A49963; A45106; A43932; B33532; A61257; P00328; P00329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002; PMID:8300571
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:121998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075; PMID:1400449
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:gl86395; PIDN:AAA59163.1; PID:gl86396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:gl86397; PIDN:AAA59164.1; PID:gl86398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)

R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:gl88863; PIDN:AAA59875.1; PID:gl88864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A;Reference number: A33532; MUID:89197956; PMID:2703501
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:gl88873; PIDN:AAA36334.1; PID:gl88874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481; PMID:1985113
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: P00328; MUID:92198477; PMID:1550588
A;Accession: P00328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: P00329
A;Molecule type: protein
A;Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:I20203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C;Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>

Alignment Scores:
Pred. No.: 7,62e-07 Length: 3020
Score: 203.00 Matches: 74
Percent Similarity: 37.90% Conservative: 20
Best Local Similarity: 29.84% Mismatches: 104
Query Match: 16.41% Indels: 50
DB: 2 Gaps: 9

US-10-081-864-14 (1-678) x A43932 (1-3020)

QY 5 CCTCTCGGAACGTCATCCCGAGTTCATGCGCTTCAAGTGGCATGGAGGCGCACCG 64
Db 1518 ProProThrThrThrProSerProProThrThrThrThr-----ThrProPro 1534
QY 65 TGAACGGCCACGAGTTCGAGATCGAGGCGAGGCGGCGCCCTACGAGGCGCACCA 124
Db 1535 ProThrThrThrProSerProProThrThrThrThrProThrThrThrThrThr 1554
QY 125 ACA---CCGTGAAGTTGAAGGTGACCAAGGCGCGCCCTCGCTTGGCGGACATCC 181
Db 1555 ThrLeuPro-----ProThrThrThrProSerProProThrThrThrThr 1569
QY 182 TGTCCTCCCGAGTCCAGTACGCTCCAGGTGACGAGGTAGTACGAGCACCCCGGACATCCCG 241
Db 1570 ThrThrProProProThrThrThrThrThrThrThrThrThrThrThrThrThr 1589

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A:Note: KIAA0324

Alignment Scores:
Pred. No.: 2,32e-05 Length: 1791
Score: 182.00 Matches: 87
Percent Similarity: 41.42% Conservative: 12
Best Local Similarity: 36.40% Mismatches: 106
Query Match: 14.56% Indels: 34
DB: 2 Gaps: 11

US-10-081-864-14 (1-678) x T02345 (1-1791)
QY 655 CTCTGGTGGCTGCTACTCTCCAGGAGTGTAGTCTCTGTTGGAGGATGATGCA 596
DB 783 ProArgThrLysThrThrSerArgArgGlyArgSerProSerProLys-----Pro 799
QY 595 GCTTGTGTCCAGTGTAGTACCGGCGAGCTGCACGGGCTCTTGGCCATGTAGATGG 536
DB 800 ArgGlyLeuGlnArgSerArgSerArgSerArgGlnLysThr-----ArgThr 816
QY 535 ACTTGAACCTCCACAGTGTAGTGGCCGCTCTCTCAGCTTCA-----GGG 491
DB 817 ThrArgArgAspArgSerGlySerSerGlnSerThrSerArgArgGlnArgSer 836
QY 490 CCTTGTGGATCTCGCCCTTCAGACCGCTCGGGGGGTACAGGCGCTCGTGGAGGCT 431
DB 837 ArgSerArgSerArgValThrArgArgArgGlyGlySerGlyThrHisSerArgSer 856
QY 430 CCACAGCCCATGTTCTTCTCT--GCATCAGGGGCGCTCGAGGGGAATTCAGCC--- 377
DB 857 ProAlaArgGlnLysSerArgThrSerArgArgArgArgGlyArgSerArgThr 876
QY 376 -----CGATGAACCTTCACTTGTAGTGAAGCAGCGCTCTCGAGGAGGAGT 329
DB 877 ProProThrSerArgLysArgSerArgSerArgThrSerProAlaProThrLysArgSer 896
QY 328 CTGGGTTCAGGTCGCGCCCGCTCTCGAGTTCATCAGCGCTCCCACTTGAAGC 269
DB 897 ArgSerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuLys 916
QY 268 -----CCTCGGGAAGGACAGCTTCTTGTAGTCGGGGATGTCGGGGGGTGTCTCA 218
DB 917 ArgArgArgSerArgSerArgThrSerProValSerArg-----ArgArgSerArgSer 934
QY 217 CGTACACCTTGGAGCGTACTGGAACCTGGGGGACAGGATGTC-----AGGCGAAG 164
DB 935 ArgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArg 954
QY 163 GCAGGGGGCGG-----CCTTGGTCACTTCACTTCACTTCACTTCACTTCACTTCA 116
DB 955 SerArgSerArgThrProProValThrArgArgArgSerArgSerArgThrProThr 974
QY 115 ---CGTAGGGGCGGCGCTCGCCCTCGCCCTCGANCTCGAATCGTGGCGGCTTCAGCGTGC 59
DB 975 ArgArgArgSerArgSerArgThrProProValThrArgArgArgSerArgSerArgThr 994
QY 58 CTTCCATGCCACCTTGAAGCGCATGNACTCGGTGTAGCTGCTCGGAGGAGGCA 2
DB 995 ProProValThr-----ArgArgArgSerArgSerArgThrSerPro 1008

RESULT 6
S53363
mucin SAC (clone JER58) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 08-Oct-1999
C:Accession: S53363

R:Guyonnet-Duperat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Gallegue-
Biochem. J. 305, 211-219, 1995
A>Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich doma
A:Reference number: S53361; MUID:95126907; PMID:7826332
A:Accession: S53363
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-279 <GUY>
A:Cross-references: EMBL:Z34278; NID:g563376; PIDN:CAA84032.1; PID:g563377
A:Experimental source: clone JER58
C:Genetics:
A:Gene: GDB:MUC5AC
A:Cross-references: GDB:454136; OMIM:158373
A:Map position: lip15.5-lip15.5
C:Keywords: Glycoprotein; tandem repeat

Alignment Scores:
Pred. No.: 5.45e-05 Length: 279
Score: 175.50 Matches: 80
Percent Similarity: 42.35% Conservative: 28
Best Local Similarity: 31.37% Mismatches: 90
Query Match: 14.19% Indels: 57
DB: 2 Gaps: 12

US-10-081-864-14 (1-678) x S53363 (1-279)
QY 5 CTCTCTCGAGAACGTATCATCCAGGATTCATGCGCTTCAAGTCCGATCGAGGGCACCG 64
DB 18 ProValProThrThrSerThrSer-----AlaPro 28
QY 65 TGAACGGCCAGAGTTCGAGATCGAGGCGAGGCGGCGCCCTACGAGGGCCACA 124
DB 29 ThrThrArgThrThrSerAlaProLysSerSerThrThrSerAlaAlaThrThrSerThr 48
QY 125 ACACCGTGAAGTGAAGGTGACCAAGGGCGGCGCCCT-----TGCCCT 166
DB 49 ThrSerGlyProGluThrThrProArgProValProThrThrSerThrThrSerPro 68
QY 167 TCGCTCGGACATCTCTCCCGCCAGTCCAGTACGCTCCAGGTGTACGTGAAGCAC 226
DB 69 ThrThrSerThrThrSerAlaProThrThrSerThrThrSerAlaSerThrThrSerThr 88
QY 227 -----CCGCGGACATCC-----CCG 241
DB 89 ThrSerGlyAlaGlyThrThrProSerProValProThrThrSerThrThrSerAlaPro 108
QY 242 ACTCAGAACACTGCTCTCCCGAGGCTTCAAGTGGGAGCGGTGATGACATTCGAGG 301
DB 109 ThrThrArgThrThrSerAlaProLysSerThrThrSerAlaAlaThrThrSerThr 128
QY 302 ACGGCGGCTGCGCAGCGTACCGAGGACTCTCTCCCTGCAGGCGGTGCTTCTATACA 361
DB 129 ThrSerGly-----ProGlyThrThrProSerProValProThrThrSerThr 144
QY 362 AGGTGAAGTTCATCGGGGTGAACCTTCCCTCCGACGCGCCCGTGTATCGAAGAAGACCA 421
DB 145 -----ThrSerAlaProThrThrSerThrThrSerGlyPro----- 156
QY 422 TGGGCTGGAGGCGCTCCACCGAGCGCTGTATCCCGCGGCGGCGGTGTGTGAAGGCGGAGA 481
DB 157 -----GlyThrThrProSerAlaValProThrThrSerThrThrSerValProThrThr 174
QY 482 TCCCAAGG-----CCCTGAAGCTGAAGGACGCGCGCCACTACTCGTGGAGT---TCA 532
DB 175 SerThrAenSerAlaProLysSerSerThrThrSerAlaAlaThrThrSerArgLysSer 194
QY 533 AGTCATCTACATGCCCA-----AGAAAGCCGTCGACGCTCGCGGCTACT 577
DB 195 GlyProGluThrThrProSerProValProThrAlaSerThrThrSerAlaSerThrThr 214
QY 578 ACTAGTGGACACCAAGCTGGACATCACT-----CCCAACAGGAGCT----- 622
DB 215 SerThrThrSerGlyProGlyThrThrProSerProValProThrThrSerThrThrSer 234

QY 623 ACACCATGTGGAGCAGTACGAGCGACCGAGGCGCCGCCACC 667
 Db 235 ValProThrThrSerThrThrSerAlaSerThrThrThrThr 249
 RESULT 7
 QMBE3
 BHLF1 protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
 C:Accession: A03742
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrall, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6092825
 A:Accession: A03742
 A:Molecule type: DNA
 A:Residues: 1-660 <BAN>
 R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
 C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
 Pred. No.: 6.57e-05 Length: 660
 Score: 175.00 Matches: 77
 Percent Similarity: 33.22% Conservative: 17
 Best Local Similarity: 27.21% Mismatches: 83
 Query Match: 14.00% Indels: 106
 Gaps: 1
 US-10-081-864-14 (1-678) x QMBE3 (1-660)

QY 673 GGAACAGGTGTGGCGGC----- 656
 Db 199 GlyProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySer 218
 QY 655 -----CTCGGTGGCTCGTACGTCTCCAGATGGTGTAGTCTCTGT--- 614
 Db 219 GlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeu 238
 QY 613 -----TGTGGAGGTGATGTCAGCT--- 593
 Db 239 ProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSerGly 258
 QY 592 -----TGGTGTCACGTAGTAGT----- 575
 Db 259 AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPro 278
 QY 574 -----ACCGGCGAGCTGCACGGCTCTTGGCCATGTAGATCG 536
 Db 279 GlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisPro----- 295
 QY 535 ACTTGAACCTCCAGAGTAGTGGCGGCTCTTACAGCTTCAGGGCCTTGTGATCTCGC 476
 Db 296 -----ProProGlyAlaGlyGlnArgPro-----SerGlyProThrGlyGlyArg 310
 QY 475 CCTTCACGACCGCTCGCGGGGTACAGGGCTCGGTGGAGGCTCCAGCCCATGGTCT 416
 Db 311 ProAlaAlaProGlyAlaProGlyThrProAla-----AlaProGlyProGlyGly 327
 QY 415 TCTTCTCATCAGCGGGCGCTCGGAGGGAAGTTTCACGCCGATGAACCTTCACCTGTAGA 356
 Db 328 GlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAla--- 346
 QY 355 TGAAGCAGCGCTCTGCAGGAGGAGTCTCTGGGTACGGTCCGCCAGCGCGCTCTCGA 296
 Db 347 ---AspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGln 365
 QY 295 -----AGTTTCATCAGCGCTCCCACTTGAAGCCCTCGGGAAGGACGAGCTTCTGT 245

Db 365 AspLeuAlaAlaAlaGlnArgCysProAlaGlyProProThrArgSerGlyAlaAla 385
 QY 244 AGTCGG-----GGATGTCGGCGGTCTTCACTGACGTACACCTTGG 206
 Db 386 AlaGlnArgThrHisArgArgProProGlyCysProArgSerAla---ArgAsnPro--- 403
 QY 205 AGCCGTACTGGAACCTGGGGGACAGGATGTCCTCCAGGGGAGGAGGGGGCGCCCTTGG 146
 Db 404 -----GlyCysPro-----ArgThrTyr 409
 QY 145 TCACCTTCACTTCACGGTGTGTGGCCCTCGTAGGGCGCCCTCGCCCT----- 95
 Db 410 -----ArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGly 425
 QY 94 ---CGCCCTCGATCTCGAACTCGTGGCGGCTTCACGGTCCCT-----CCA 53
 Db 426 GlnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrPro 445
 QY 52 TCCGCACCT 44
 Db 446 AlaAlaPro 448
 RESULT 8
 S55316
 mucin (clone PGM-2B) - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 17-Nov-2000
 C:Accession: S55316
 R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulos-Cladaras, M.; Specian, R.D.; LaMont, J.T.
 Biochem. J. 308, 89-96, 1995
 A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
 A:Reference number: S55315; MUID:95275264; PMID:7755593
 A:Accession: S55316
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-317 <TUR>
 A:Cross-references: GB:U12768; NID:915204; PIDN:AAC48525.1; PID:915207
 C:Superfamily: pig submaxillary mucin

Alignment Scores:
 Pred. No.: 7.1e-05 Length: 317
 Score: 174.00 Matches: 81
 Percent Similarity: 45.23% Conservative: 28
 Best Local Similarity: 33.61% Mismatches: 92
 Query Match: 14.07% Indels: 40
 Gaps: 2
 US-10-081-864-14 (1-678) x S55316 (1-317)
 QY 8 CTTCCGAGAACGTCTATCA-----CCGAGTTCA 34
 Db 51 ProProGluThrSerSerHisGlyAlaThrSerSerThrThrSerValGlnProSerSer 70
 QY 35 TGCCTTCAGTCCGATGAGGGCACCGTGAACGCCACGATTCGATCGAGGGCG 94
 Db 71 SerSerSer-----AlaProThrThrSerAlaThrSerValGlnProSer 85
 QY 95 AGGCGAGGGCGCCCTTACGAGGGCCACACCGTGAAGTTGAAGTGACCAAGGGCG 154
 Db 86 SerSerGlySerAlaProThrThrSerAlaThrSerValGlnSerSerSerGlySer 105
 QY 155 GCCCCTGCTCCCTCGCTGGGACATCTGTCCTCCAGTTCAGTACGGTCTCA----- 208
 Db 106 AlaProThrThrSer---AlaThrSerValGlnProSerSerSerSerProProle 124
 QY 209 AGGTGTACGTGAACACCCCGCAGCATCC-----CCGACTACAAGAGCTGT 256
 Db 125 SerSerThrThrSerValGlnProSerSerSerSerSerAlaProThrThrSerAlaThr 144
 QY 257 CTTTCCCGAGGGCTTCAAGTGGAGCGCTGATGAATTCGAGGACGGCGCGTGGCGA 316
 Db 145 SerValGlnSerSerSerSerGlySerAlaProThrThrSerAlaThrSerValGlnPro 164

antifreeze glycopeptide AFGP polypeptide precursor [imported] - Boreogadus saida
C:Species: Boreogadus saida
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C/Accession: T44768
R:Chen, L.; Devries, A.L.; Cheng, C.H.C.
Proc. Natl. Acad. Sci. U.S.A. 94, 3817-3822, 1997

A>Title: Convergent evolution of antifreeze glycopeptides in Antarctic notothenioid fish
A/Reference number: Z22834; MUID:97268653; PMID:9108061
A/Accession: T44768
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 1-507 <CHE>
A/Cross-references: EMBL:U43200; NID:92078482; PIDN:AAC60129.1; PID:92078483

C/Genetics:
A/Introns: 1/3

Alignment Scores:
Pred. No.: 0.000465 Length: 507
Score: 163.00 Matches: 66
Percent Similarity: 43.64% Conservative: 37
Best Local Similarity: 27.97% Mismatches: 77
Query Match: 13.18% Indels: 56
DB: 2 Gaps: 11

US-10-081-864-14 (1-678) x T44768 (1-507)

```
QY 8 CTCCTCGAGAACCTCATCCAGTTCATCGCTTCAAGTGGCGATGAGGCGCCGTGA 67
DB 128 ProAlaArgAlaAlaThrProAlaThrAlaAlaThr-----ProAla 141
QY 68 AGCGCCACGAGTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGCGCCACAACA 127
DB 142 ThrAlaAlaThrProAlaThrAlaAlaThrAlaAlaThrAlaAlaThrAla 161
QY 128 CGGTGAAGTTGAAGTGTGACCAAGGCGCGCCCTCGCTTCCCTGGGACATCCTGTC 187
DB 162 -----AlaThrAlaAlaArgAlaAlaThrProAlaThrAlaAlaThr 175
QY 188 CCCAGTTC-----AGTACGCTCCAGTGTACGTGAGCACCCTCCGACATCCCGACT 244
DB 176 ProAlaThrProAlaThrAlaAlaArgAlaAlaThrProAlaThrAlaAla 195
QY 245 ACAAGAAGGTGTCTCTCCCGAGGCGTTCAGTGGGCGCGTGTGATGAATTCGAGGACG 304
DB 196 ThrAlaAlaThrAlaAlaThrAlaAlaThrAlaAlaThrAlaAlaThr 214
QY 305 GCGGCGTGGCGACCGGTGACCCAGGACTCTCCTCGTGGAGCGGCTCTCATCTACAGG 364
DB 215 AlaAla-----ThrProAlaArgAlaAlaAlaAlaThr--- 226
QY 365 TGAAGTTCATCGCGGTGAATCTCCCTCCGAGCGCCCGTGTGATGCAGAGAAGACCATGG 424
DB 227 -----ProAlaThrAlaPro----- 231
QY 425 GCTGGGAGGCTCCACCGAGCGCTGTACCCCGAGCGCGTGTGAAG----- 475
DB 232 -----ThrProAlaThrAlaAlaThrProAlaThrAlaAlaThrAlaAla 248
QY 476 -----GGGAGATCCACAGCCCTGAAGCTGAGGAGCGGCGGCGCTACT 517
DB 249 ProThrAlaAlaThrProAlaArgAlaAlaArgAlaAlaThrProAlaThrAlaAlaThr 268
QY 518 ACCTGGTGGAGTTCAGATCCATCTACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACT 577
DB 269 LeuAlaThrAlaAlaThrProAlaThr---ProAlaThrProAlaThr-----AlaAla 285
QY 578 ACTACGTGGACACCAAGCTGGATCATCCTCCCAACACGAGGACTACATCTGTGGAGC 637
DB 286 ThrAspAlaThrAlaAlaThrAlaAlaThrProAlaArgAlaAlaThrProAlaThrPro 305
QY 638 AGTACGAGCGCA-----CCGAGGCGCGCCACCTGTTC 673
DB 306 AlaThrAlaAlaThrProAlaThrProAlaThrAlaAlaThrAlaAla 321
```

Search completed: July 29, 2004, 14:37:30
Job time : 23.6859 secs

Blank Sheet USPTO

OM nucleic - protein search, using frame_plus_n2p model
Run on: July 29, 2004, 14:24:20 ; Search time 9.79061 Seconds
(without alignments)
7211.716 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 agggcctctcgagacgt.....gcaccacctgtctctgtaa 678

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2_1/USPTO.spool.p/US10081864/runat 29072004 150505 25463/app query.fasta_1.1742
-DB=SwissProt 42 -QPMF=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10081864 @CNG 1.1 22 @runat 29072004 150505 25463 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224.5	18.1	238	1	GFP_AEQVI
2	203	16.4	5179	1	MUC2_HUMAN
3	175	14.0	660	1	YHL1_EBV
4	170.5	13.8	1367	1	AMYL_YEAST
5	165.5	13.4	1161	1	DAN4_YEAST
6	161	12.9	1033	1	IP2_STRCO
7	159.5	12.9	267	1	EXTN1_MAIZE
8	153.5	12.4	283	1	EXTN1_SORBI
9	151.5	12.2	797	1	VGLX_HSVB
10	150.5	12.2	5703	1	MUSB_HUMAN
11	150	12.1	707	1	SFPQ_HUMAN
12	150	12.0	1046	1	IP2_STRAW
13	148.5	12.0	1513	1	MUC2_RAT
14	148	11.8	659	1	SFRG_HUMAN
15	146	11.8	1233	1	MUSA_HUMAN
16	145.5	11.8	437	1	YGY3_HALSQ
17	145.5	11.8	573	1	C114_MOUSE
18	145.5	11.8	1336	1	W146_HUMAN

19	144.5	11.7	307	1	SGS3_DROME
20	144.5	11.6	1461	1	IE18_PRVIF
21	144	11.6	600	1	SP96_DICDI
22	143.5	11.5	348	1	YT35_STRFR
23	143.5	11.5	497	1	VE2_HPV20
24	142	11.4	653	1	SFRG_MOUSE
25	142	11.4	825	1	ICPO_HSV2H
26	141.5	11.4	297	1	SGS4_DROME
27	141.5	11.4	555	1	GP1_CHLRE
28	141	11.4	217	1	SGS3_DROSI
29	140.5	11.2	514	1	VE2_HPV5B
30	139.5	11.3	662	1	MUC1_XENLA
31	139.5	11.3	699	1	VGLG_HSV2H
32	139.5	11.2	825	1	SE5_RAT
33	139	11.1	1446	1	IE18_PRVKA
34	138	11.2	1509	1	GSRI_HUMAN
35	137.5	11.1	349	1	ENL2_ARATH
36	137.5	11.1	1733	1	VNUA_PRVKA
37	137	11.0	502	1	VE2_HPV25
38	137	11.0	1226	1	CDN1_HUMAN
39	136	11.0	261	1	PRP2_MOUSE
40	136	10.9	514	1	VE2_HPV05
41	136	10.9	2424	1	CCAA_RABIT
42	136	11.0	3149	1	TEGU_EBV
43	135.5	10.8	431	1	UL61_HCMVA
44	135	10.9	263	1	SGS3_DROYA
45	133.5	10.7	512	1	IE63_HSV2H

ALIGNMENTS

RESULT 1	GFP_AEQVI	STANDARD;	PRT;	238 AA.
ID	GFP_AEQVI	STANDARD;		
AC	P4212; Q17104; Q27903;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;			
OC	Aequorea; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RP	MEDLINE=9217527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cornier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein."			
RL	Gene 111:229-233 (1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein."			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=97299832; PubMed=9154981;			
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;			
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent			
RT	protein by modification of its codon usage."			
RL	Plant Mol. Biol. 33:989-999(1997).			
RN	[4]			
RP	CHROMOPHORE.			
RP	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RT	green-fluorescent protein."			
RL	Biochemistry 32:1212-1218(1993).			

Best local Similarity: 25.32% Mismatches: 111
 Query Match: 18.15% Indels: 13
 DB: 1 Gaps: 6

US-10-081-864-14 (1-678) x GFP_AEQVI (1-238)

QY	1	ATGGCTCTCTCCGAGAACGTCATCACCAGAGTTTCATCGCTTCAAGTGGCGCATGAGGCG	60
Db	1	MetSerlysglygluGluLeuPheThrGlyValProIleuValGluLeuAspGly	20
QY	61	ACCGTGAACGCCACGAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGCG	120
Db	21	AspValAsnGlyHisLysPheSerValserGlyGluGlyGluAspAlaThrTyrGly	40
QY	121	CACACACCGTGAAGTGAAGTGACCAAGGGGGCGCCCTCCCTCCCTGGGACATC	180
Db	41	LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThrProThr	59
QY	181	CTGTCCCCCAGTTCACGATACGGCTCCAAAGGTGACGTGAAGCACACCCCGCACATCCCC	240
Db	60	LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys	79
QY	241	-----GACTACAAAGAGTGTCTCTCCGAGGGCTTCAAGTGGGAGCGCGTGAAC	294
Db	80	GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe	99
QY	295	TTTCGAGGACGGCGGTGGCGACCGTGCACCCAGAGACTCTCCCTCGCAGCAGCGCTGCTC	354
Db	100	PhelysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu	119
QY	355	ATCTACAAGGTGAAGTTCATCGCGGTGAATTCCTCCCGACCGGCCCGGTGATCGAAG	414
Db	120	ValAsnArgIleGluLeuLysGlyIleAspPhelysGluAspGlyAsnIleLeuGlyHis	139
QY	415	AGACCATGGGTGGGAGCGCTCCACCGAGCGCGTGTACCCCGCAGCGCGGTGCTGAAG	474
Db	140	Lys---LeuGluTyrAsnTyrAsnSerHisasnValTyrIleMetalaspiysGlnLys	158
QY	475	CGCGAGATCCACAGCGCCTGAAGTCTGAAG-----GACCGCGGCCCACTACCTG	522
Db	159	AsnGlyIleLysValAsnPhelysIleargHisasnIleGluAspGlySerValGlnLeu	178
QY	523	GTGGAGTTC-----AATGTCATCTACATGCCAAGAGCGCGTGCAGTGTCCCGGCTAC	576
Db	179	AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn	198
QY	577	TACTAGTGGACACCAAGCTGGACATCACC-----TCCCAACAGGAGACTACACC	627
Db	199	HsTyrLeuSerThrGlnSerAlaLeuSerLysAspProasnGluLysArgAspHisMet	218
QY	628	ATCTGTGGAGCATACGAGCGCACCGACCGCGCGCCGCCACCATC	666
Db	219	ValLeuLeuGluPheValThrAlaAlaGlyIleThrHis	231

RESULT 2

MUC2_HUMAN	STANDARD;	PRT;	5179 AA.
ID	MUC2_HUMAN	Q02817; Q14878;	
AC	Q02817; Q14878;	01-JUN-1994 (Rel. 29, Created)	
DT	DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Mucin 2 precursor (Intestinal mucin 2).		
GN	MUC2 OR SMUC.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	Tissue=Intestine;		
RX	MEDLINE=94132002; PubMed=8300571;		
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;		
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.		

RT identification of the amino terminus and overall sequence similarity
 RT to prepro-von Willebrand factor.";
 RT J. Biol. Chem. 269:2440-2446(1994).
 RN [2]
 RN SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
 RN TISSUE=Colon;
 RN MEDLINE=93016075; PubMed=1400449;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
 RA Kim Y.S.;
 RA "The human MUC2 intestinal mucin has cysteine-rich subdomains located
 RT both upstream and downstream of its central repetitive region.";
 RT J. Biol. Chem. 267:21375-21383(1992).
 RN [3]
 RN SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
 RN MEDLINE=91598717; PubMed=1885763;
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
 RA Petersen G.M., Kim Y.S.;
 RA "MUC-2 human small intestinal mucin gene structure. Repeated arrays
 RT and polymorphism.";
 RT J. Clin. Invest. 88:1005-1013(1991).
 CC -I- FUNCTION: Coats the epithelia of the intestines, airways, and
 CC other mucus membrane-containing organs. Thought to provide a
 CC protective, lubricating barrier against particles and infectious
 CC agents at mucosal surfaces.
 CC -I- SUBUNIT: Multimeric.
 CC -I- SUBCELLULAR LOCATION: Secreted.
 CC -I- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
 CC BRONCHUS, CERVIX AND GALL BLADDER.
 CC -I- PM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
 CC -I- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -I- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -I- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -I- SIMILARITY: Contains 2 VWFC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMEL; L21998; AAB95295.1; -;
 DR EMEL; M74027; AAA59875.1; -;
 DR EMEL; M94131; AAA59163.1; -;
 DR EMEL; M94132; AAA59164.1; -;
 DR PIR; A49963; A49332.
 DR Genew; HGNC:7512; MUC2.
 DR MIM; 158370; -;
 DR GO; GO:0005803; C:secretory vesicle; TAS.
 DR InterPro; IPR006208; Cys_knot.
 DR InterPro; IPR006207; Cys_knot_C.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002919; TIL_Cysrich.
 DR InterPro; IPR001007; VWFC_C.
 DR InterPro; IPR001846; VWFC_D.
 DR Pfam; PF00007; Cys_knot; 1.
 DR Pfam; PF01826; TIL; 1.
 DR Pfam; PF00093; vwc; 1.
 DR Pfam; PF00094; vwd; 4.
 DR SMART; SMO0214; VWC; 2.
 DR SMART; SMO0216; VWD; 4.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS01208; VWFC_1; 2.
 DR PROSITE; PS0184; VWFC_2; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 5179 MUCIN 2.
 FT DOMAIN 1401 1747 APPROPRIATE REPEATS.

```
FT REPEAT 1401 1416
FT REPEAT 1417 1432
FT REPEAT 1433 1448
FT REPEAT 1449 1464
FT REPEAT 1465 1471
FT REPEAT 1472 1478
FT REPEAT 1479 1494
FT REPEAT 1495 1517
FT REPEAT 1518 1533
FT REPEAT 1534 1556
FT REPEAT 1557 1572
FT REPEAT 1573 1596
FT REPEAT 1597 1612
FT REPEAT 1613 1635
FT REPEAT 1636 1651
FT REPEAT 1652 1675
FT REPEAT 1676 1683
FT REPEAT 1684 1699
FT REPEAT 1700 1715
FT REPEAT 1716 1731
FT REPEAT 1732 1747
FT DOMAIN 4815 4886
FT DOMAIN 4924 4991
FT DOMAIN 5075 5160
FT DISULFID 5075 5122
FT DISULFID 5089 5136
FT DISULFID 5098 5152
FT DISULFID 5102 5154
FT DISULFID ? 5159
FT CARBOHYD 163 163
FT CARBOHYD 423 423
FT CARBOHYD 670 670
FT CARBOHYD 770 770
FT CARBOHYD 894 894
FT CARBOHYD 1139 1139
FT CARBOHYD 1154 1154
FT CARBOHYD 1215 1215
FT CARBOHYD 1230 1230
FT CARBOHYD 1246 1246
FT CARBOHYD 1787 1787
FT CARBOHYD 1820 1820
FT CARBOHYD 4339 4339
FT CARBOHYD 4351 4351
FT CARBOHYD 4362 4362
FT CARBOHYD 4373 4373
FT CARBOHYD 4422 4422
FT CARBOHYD 4438 4438
FT CARBOHYD 4502 4502
FT CARBOHYD 4616 4616
FT CARBOHYD 4627 4627
FT CARBOHYD 4752 4752
FT CARBOHYD 4787 4787
FT CARBOHYD 4881 4881
FT CARBOHYD 4888 4888
FT CARBOHYD 4955 4955
FT CARBOHYD 4970 4970
FT CARBOHYD 5019 5019
FT CARBOHYD 5038 5038
FT CARBOHYD 5069 5069
FT CARBOHYD 1351 1351
FT CONFLICT 1412 1412
FT CONFLICT 1449 1449
FT CONFLICT 1504 1504
FT CONFLICT 4192 4192
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
```

Alignment Scores:

```
Pred. No.: 1.57e-05
Score: 203.00
Percent Similarity: 37.90%
Best Local Similarity: 29.84%
Query Match: 16.41%
DB: 1
Matches: 74
Conservatives: 20
Mismatches: 104
Indels: 50
Gaps: 9
```

```
US-10-081-864-14 (1-678) x MUC2_HUMAN (1-5179)
QY 5 CCTCTCCGAGAACGTCATCACCGAGTTTCATCGGCTTCAAGGTGCGCATGAGGCGCACCG 64
Db 1518 ProProThrThrThrProSerProThrThrThrThrThrThrThrThrThrThrThrThr
QY 65 TGAAGCGGCACGAGTTCGAGATCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 124
Db 1535 ProThrThrThrProSerProProThrThrThrThrThrThrThrThrThrThrThrThr
QY 125 ACA---CGTGAAGTTGAAGGTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181
Db 1555 ThrLeuPro-----ProThrThrThrProSerProProThrThrThrThrThrThrThr
QY 182 TGTCCCGCCAGTTCAGTACGCTCCCAAGGTGAGTGAAGCAGCCCGCCGACATCCCGC 241
Db 1570 ThrThrProProProThrThrThrProSerProProThrThrThrThrThrThrThrThr
QY 242 ACTACAAGAAGCTGCTCTCCCGAGGCTTCAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 301
Db 1590 ThrThrThrThrThrThrProProThrThrThrThrThrThrThrThrThrThrThrThr
QY 302 ACGGCGGCGTGGGACCGTGAACCCAGGACTCTCCCTGCGAGGCGGCGGCGGCGGCGGCGG 361
Db 1610 ThrThr-----ProProProThrThrThrThrThrThrThrThrThrThrThrThr
QY 362 AGGTGAAGTTCATCGGCGTGAACCTCCCTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421
Db 1627 ThrProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr
QY 422 TGGGCTGGGAGGCGCTCCACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 472
Db 1640 ---ThrProSerProProProThrThrThrThrThrThrThrThrThrThrThrThrThr
QY 473 AGGCGGAGATCCACAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 532
Db 1659 ProProThrThrThrThrThrPro-----SerProProThrThrThrThrThrThrThr
QY 533 AGTCCATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 562
Db 1676 ProProThrThrThrThrThrProSerProThrThrThrThrThrThrThrThrThrThr
QY 563 -----AGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 592
Db 1696 MetThrThrProSerProThrThrThrProSerProThrThrThrThrThrThrThrThr
QY 593 AGCTGGACATCA-----CCTCCCAACAGGAGTACACCATCGTGGAGGAGT 640
Db 1716 SerSerThrThrThrProSerProProProThrThrThrThrThrThrThrThrThrThr
QY 641 ACGAGCGCACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 664
Db 1736 ThrProSerProProThrThrThrThr 1743

RESULT 3
YHLL1_EBV
ID YHLL1_EBV STANDARD; PRT; 660 AA.
AC F03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLF1 protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hutchinson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
```

"DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211(1984).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; V01555; ; NOT_ANNOTATED_CDS.
 DR FIR; A03742; Q0B53.
 KW Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;
 Alignment Scores: Pred. No.: 0.000778 Length: 660
 Score: 175.00 Matches: 77
 Percent Similarity: 33.22% Conservative: 17
 Best Local Similarity: 27.21% Mismatches: 83
 Query Match: 14.00% Indels: 106
 DB: Gaps: 16

US-10-081-864-14 (1-678) x YHL1_EBV (1-660)

QY	673	GGACAGGTGGTGGCGG	-----	656
Db	199	GlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySer	218	
QY	655	-----CCTCGGTGGCTGCTGCTCCAGTGGTGTAGTCTCTGCT	614	
Db	219	GlyProAlaAspProProAlaAlaAlaArgLeuProGluArgGlnGluProArgLeu	238	
QY	613	-----TGTGGAGGTGATGTCAGCT	593	
Db	239	ProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProProThrArgSerGly	258	
QY	592	-----TGGTGCCAGTGTAGT	575	
Db	259	AlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPro	278	
QY	574	-----AGCGGGCAGCTGCACGGGCTTCTTGGCCATGTAGATGG	536	
Db	279	GlyCysProArgThrTrpArgArgSerGlyAlaGlnArgGlyHisPro	295	
QY	535	ACTTGAACCTCCACAGTAGTGGCGCGCTTCAGCTTCAGGCGCTTGTGGATCTCG	476	
Db	296	-----ProProGlyAlaGlyGlnArgPro-----SerGlyProThrGlyGlyArg	310	
QY	475	CCTTACGACGCGCTCGCGGGGTACAGGCGCTCGGTGGAGGCTCCAGCCCATGCTCT	416	
Db	311	ProAlaAlaProGlyAlaProGlyThrProAla-----AlaProGlyProGlyGly	327	
QY	415	TCTTCTGCATCAGGGCGCTCGAGGGAGTTCACCGCATGAATTCACCTTCGTAGA	356	
Db	328	GlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAla---	346	
QY	355	TGAAGCAGCGCTCTCGCAGGAGGAGTCTTGGGTGCGGCGCGCCCTCTCGA	296	
Db	347	---AspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGln	365	
QY	295	-----AGTTCATCAGCGCTCCACCTTGAAGCCCTCGGGAGGACAGCTTCTGT	245	
Db	366	AspLeuAlaAlaGlnArgCysProAlaGlyProProThrArgSerGlyAlaAla	385	
QY	244	AGTCGG-----GGATGTCGGGGGTGCTTCACGTACACCTTGG	206	

Db	386	AlaGlnArgThrHisArgArgProProGlyCysProArgSerAla---	403
QY	205	AGCGGTACTGAACTGGGGGACAGGATGTCCAGGCGAAGGAGGGGCGCCCTTGG	146
Db	404	-----GlyCysPro-----ArgThrTrp	409
QY	145	TCACCTTCAACTTCACGGTGTGTGGCCCTCTAGGGCGGCGCTCGCCCT	95
Db	410	-----ArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGly	425
QY	94	---GCGCCTCGATCTCGAACTCGTGGCCCTTCAGGTCAGGTCGCT	53
Db	426	GlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrPro	445
QY	52	TGCGCACCT 44	
Db	446	AlaAlaPro 448	

RESULT 4
 AMVH_YEAST
 ID AMVH_YEAST STANDARD; PRT; 1367 AA.
 AC P08640; P08068;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Glucosylase 1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 GN STAL OR STA2 OR MAU5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313266; PubMed=9169870;
 RA Churcher C.M., Bowman S., Baddock K., Bankier A., Brown D., Chillingworth T., Connor R., Delvin K., Gentles S., Hamlin N., Harris D.E., Horenell T., Hunt S., Jagels K., Jones M., Lye G., Moulton J., Odell C., Pearson D., Rajadream M.A., Rice P., Rowley N., Skelton J., Smith V., Walsh S., Whitehead M.A., Barrell B.G.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
 RT Nature 387:84-87 (1997).
 RL [2]
 RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RP MEDLINE=87194500; PubMed=3106330;
 RX Yamashita I., Nakamura M., Fukui S.;
 RA "Gene fusion is a possible mechanism underlying the evolution of STAL.";
 RT J. Bacteriol. 169:2142-2149 (1987).
 RL [3]
 RN SEQUENCE OF 1-31 FROM N.A.
 RP STRAIN=SPX101-1C;
 RX MEDLINE=89031230; PubMed=3141213;
 RA Pardo J.M., Inez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae.";
 RL FEBS Lett. 239:179-184 (1988).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.
 CC -!- SIMILARITY: TO S.POMBE SPBC215.13.
 CC -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; Z38061; CAA86176.1; -.


```

DB: 1 Gaps: 9
US-10-081-864-14 (1-678) x DAN4_YEAST (1-1161)
QY 17 AGCTCATCCAGGCTTCAAGTCCGATCGAGGCGACCGTGAAGCCACG 76
D 143 ThrSerThrThrSerThrThrSerThrThrThrThrThrThrThrThr 157
QY 77 AGTTGAGATCGAGGCGAGGCGAGGCGCGCCCTCAGAGGCGACACACCGTGAAGT 136
D 158 ThrSer-----ThrThrProThrThrSerThrThr----- 167
QY 137 TGAAGTGACCAAGGCGCGCCCTCGCTCGCTGGACATCTGCCCCCAGTTCC 196
D 168 -----SerThrThrProThrThrSerThrThrThrThrProThrThrSer 183
QY 197 AGTACGGCTCAAGGTGACGAGGCGAGGCGCGACATCCCGCAGTACA-----AGAAGC 253
D 184 ThrThrSerThrThrProThrThrSerThrThrThrThrThrThrThrThr 203
QY 254 TGTCCTTCCCGAGGCTTCAAGTGGGAGCGGCGTGTGAGTCTCGAGGCGCGCGTGG 313
D 204 SerThrThrProThrThrSerThrThrThrThrThrThrThrThrThrThr 223
QY 314 CGACCGTGACCCAGGACTCTCCCTCGAGGCGGCTGTCTCATCTACAGGTGAAGTTCA 373
D 224 ThrProThrThrSerThrThrProThrThrThrThrThrThrThrThrThr 243
QY 374 TCGCGGTGAACCTTCCCTCCGAGCGCCCGTGTGATGAGAGAACGACCATGGCTGGAGG 433
D 244 ThrLysSerThrThrProThrThrSer----- 252
QY 434 CCTCCACGAGCGCTGTACCCCGCGAGCGGCGTGTGAGGCGGAGATCCACAGGCC 493
D 253 -----SerThrThrThrThrThr-----SerThrAlaProThrPro 264
QY 494 TGAAGCTGAAGGAGCGGCGGCGACCTACTGCTGGTGAAGTCAAGTCCATCTACATGCCAAGA 553
D 265 -----ThrThrSerThrThr-----SerThrAlaProThrThr 275
QY 554 AGCCCGTGAGTGGCGCGCTACTAGTGTGACACCAAGCTGGACATCACTCCACCA 613
D 276 -----SerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 292
QY 614 ACGAGGATACACCATCGTGAGCAGTACGAGGCGGCGCGCGCGCCACCATCTGTTCC 673
D 293 ThrSerThrThrSerThrThrPheSerThrSerThrSerAlaSerAlaSerValIleSer 312

RESULT 6
ID IF2_STRCO STANDARD; PRT; 1033 AA.
AC Q8CJ08;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INF8 OR SC05706 OR SC9P2.10C OR SC5H4.30.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MSDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

```

```

RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
CC Nature 417:141-147(2002).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AL939124; CAD55362.1; -.
DR PIR; T35989; T35989.
DR HAMAP; MF_00100; -.
DR InterPro; IPR000637; At_hook.
DR InterPro; IPR001851; Bac_inmem_transp.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR008847; IF2_N.
DR InterPro; IPR001806; Ras_transfmg.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00009; GTP_EFTU_D2.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR Pfam; PF04760; IF2_N; 2.
DR PRINTS; PR00529; ATHOOK.
DR PRINTS; PR00315; ELONGATNFACT.
DR PRINTS; PR00449; RASTRNSFRNG.
DR ProDom; PD186100; IF2; 1.
DR TIGRfams; TIGR00487; IF-2; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 529 681 G-DOMAIN.
FT NP_BIND 535 542 GTP (BY SIMILARITY).
FT NP_BIND 585 589 GTP (BY SIMILARITY).
FT NP_BIND 639 642 GTP (BY SIMILARITY).
SQ SEQUENCE 1033 AA; 105657 MW; 47D581FB0072A045 CRC64;

Alignment Scores:
Pred. No.: 0.00588 Length: 1033
Score: 161.00 Matches: 65
Percent Similarity: 38.61% Conservative: 13
Best Local Similarity: 32.18% Mismatches: 78
Query Match: 12.88% Indels: 46
DB: 1 Gaps: 11

US-10-081-864-14 (1-678) x IF2_STRCO (1-1033)
QY 655 CCTCGTGGCGCTCGTACTGCTCCAGATGCTGTAGTCTCTGTGGAGGTGATGCCA 596
D 135 ProArgProAlaProAlaProGluPheThrAlaPro-----Pro 148
QY 595 GCTTGGTGTCCAGCTAGTAGTAGCGGGCGACGTGACGGGCTTCTTGGCCATGATAGG 536
D 149 Ala---AlaProAlaProSerThrProAlaProAlaProSerGlyPro----- 164
QY 535 ACTTGAACTCCACCGAGTAGTGGCGCGCTCTTCAGCTTCAGGCGCTTGTGGATCTCGC 476
D 165 -----LysProGly---GlyAlaArgProGlyAlaProLysProGlyGlyAlaArg 180
QY 475 CCTTCA-----GCACGCGCTGCGGGGGGTACAGCGCTCGGTGGAGGCTCCCGACCCA 422

```

```

181 ProSerGlyProGlyGlnAspArgGlyGlnGlnGlyGlnGlyArgPro----- 197
421 TGGTCTCTTCTGATCATACGGGGCGCTCGAGGGAAGTTACACCGGATGAATTCACCT 362
198 -----GlyGlnArgProGlyAlaPro----- 205
361 TGTAGTAGAAGCAGCGCTCTGAGGAGGAGTCTGGGTACGGTCCGACCGCCGCT 302
206 -----AlaGlnArgProGlyGlyArg-----ProGlyGlyProArgProGlyAsnAsn 221
301 CCTGAAAGTTCA-----TCAGCGCTCCCACTTGAAGCCCTCGGGGA 260
222 ProPheThrSerGlyGlyAsnAlaGlyMetAlaArgProSerAlaProArgProGlnGly 241
259 AGGACAGCTTCTTGTAGTCGGGATGTCGGGGGGTCTTCAGTACACCTTGGAGCGCT 200
242 GlyProArgProGlyGlyProGlyGlyAlaProGlyGlyGlyProArgProGlnGlyPro 261
199 ACTGGAAGTGGGGGACAGGATGTCACAGGCGAAGGCGAGGCGCGCCCTTGTACACT 140
262 GlyGlyGlnGlyGlyProArgProGlnAlaProGlyGlyAsnArgProSerProGly 281
139 TCAACTTCACGG-----TGTTGGGCTCTGT---AGGGCGGCGCTCGCCCT 95
282 SerMetProArgProGlnGlyGlyGlyAlaGlyProArgProGlyGlyProArgPro 301
94 CGCCCT 89
302 AsnPro 303

RESULT 7
EXTN MAIZE STANDARD; PRT; 267 AA.
AC P14918;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Extensin precursor (Proline-rich glycoprotein).
GN HRGP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NGBI_taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. W64A, and cv. E41;
RA Stiefel V., Perez-Grau L., Albericio F., Giralt E., Ruiz-Avila L.,
RA Ludevid M.D., Puigdomenech P.
RT "Molecular cloning of cDNAs encoding a putative cell wall protein from
RL Zea mays and immunological identification of related polypeptides.";
RL Plant Mol. Biol. 11:483-493(1988).
CC -!- FUNCTION: Structural component in primary cell wall.
CC -!- TISSUE SPECIFICITY: Mainly in the coleoptile node and root tip.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13499; CAA31854.1; -
DR EMBL; X13506; CAA31860.1; -
DR EMBL; M36912; AAA33455.1; -
DR EMBL; M36913; AAA33456.1; -
DR EMBL; M36914; AAA33457.1; -
DR PIR; S08314; S08314.

```

```

DR MaizeDB; 17152; -.
DR InterPro; IPR003882; Pistill_extensin.
DR PRINTS; PRO1218; PSTLTEXTENSIN.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 ?
FT CHAIN ? 267
FT DOMAIN 18 253
FT REPEAT 18 33
FT REPEAT 34 54
FT REPEAT 55 70
FT REPEAT 71 91
FT REPEAT 92 107
FT REPEAT 108 128
FT REPEAT 129 144
FT REPEAT 145 160
FT REPEAT 161 179
FT REPEAT 180 195
FT REPEAT 196 211
FT REPEAT 212 232
FT REPEAT 233 253
FT DOMAIN 261 265
FT VARIANT 245 245
FT VARIANT 261 261
SQ SEQUENCE 267 AA; 28349 MW; A6F406F4645FEECB CRC64;

Alignment Scores:
Pred. No.: 0.00684 Length: 267
Score: 159.50 Matches: 75
Percent Similarity: 39.32% Conservative: 17
Best Local Similarity: 32.05% Mismatches: 94
Query Match: 12.89% Indels: 48
DB: 1 Gaps: 14

US-10-081-864-14 (1-678) x EXTN_MAIZE (1-267)
QY 5 CCTCTCCGAGAACGTCAATCACCAGTTCATGCCCTTCAAGTCCGATCGGAGGCGACCG 64
DB 34 ProProThrTyThrProSerProLysProAlaSerLysProProThrProLysPro 53
QY 65 TGAACGGCCACGAGTTCAGATCAGGCGGAGGCGGCGCCCTACGAGGCCACA 124
DB 54 ThrProProThrTyThrProSer-----ProLysProProThrProLysPro 69
QY 125 ACACGGTGAAGTTGAAGGTACCAAGGGCGGCGCCCTGCCCTTCCGCTGGGACATCTGT 184
DB 70 ThrPro-----ProThrTyThrProSerProLysProProAlaThrLys 84
QY 185 CCCCCCAGT---TCCAGTACGGCTCCAGGTGTACGTGAAGC-----ACCCCG 229
DB 85 ProProThrProLysProProThrTyThrProSerProLysProProThrPro 104
QY 230 ---CCGACATCCCGACTACAGAAGCTGCTTCCCGAGGGCTTCAAGTGGGAGCGCG 286
DB 105 LysProThrProProThrTyThrProSerProLysProProAlaThrLys----- 121
QY 287 TGAAGAACTTCGAGGACGGCGGCTGGCGACCGGAGCCAGG---ACTCCTCCCTGCAGG 343
DB 122 -----ProProThrProLysProProThrTyThr 132
QY 344 ACGGCTGCTTCACTACAAGGTGAAGTTCATCGCGTGAACCTTCCCTCCGAGGCGCCCG 403
DB 133 ThrPro-----SerProLysProProThrProLysProProThrPro 146
QY 404 TGATGCAGAGAAGACCATGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGGAGCG 463
DB 147 ThrTyThrProSerPro-----LysProProThrProLysProProThr 163
QY 464 CGGTGCTGAAGGGCGGAGATCCACAAGGCCCTGAAGCTGAAGGAGCGGCGGCGGCTACCTGG 523
DB 164 -----TyrThrProSerProLysProProThrHisProThrProLysProProThr 180
QY 524 TGGAGTTCAAGTCCATCTACATGCCCAAGAGCCCGGTGCGAGCTGCCCGGCTACTACTAG 583

```

```

Db 181 ProThrTyrThrProSer-----ProLysProProThrProProLysProProThrProProThr 198
QY 584 TGACACCAAGTGGACATCACTCCACACAGGAGTACACCATCGGAGCAGTACG 643
Db 199 TyrThrProSer---ProLysProProThrProProLysProProThrProPro---ThrTyrThr 216
QY 644 AGGCAACGAGGCGCCCA-----CCACTGTCTCT 674
Db 217 ProSerProLys-ProProAlaThrLysProProThrPro 229

RESULT 8
EXTN_SORBI STANDARD; PRT; 283 AA.
AC P24152;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Extensin precursor (Proline-rich glycoprotein).
GN ERGP.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
OX NCBI_TaxID=4559;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=leaf;
RX MEDLINE=91370882; PubMed=1893107;
RA Raz R., Cretin C., Puigdemonech P., Martinez-Izquierdo J.A.;
RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum
vulgare.";
RL Plant Mol. Biol. 16:365-367(1991).
CC -!- FUNCTION: Structural component in primary cell wall.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
GLYCOSYLATED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; X56010; CAA39485.1; -.
DR PIR; S13383;
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR KEGG; K01218; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 283 EXTENSIN.
SQ SEQUENCE 283 AA; 29593 MW; 5D7FCD0DA8ED2D90 CRC64;

Alignment Scores:
Pred. No.: 0.0162 Length: 283
Score: 153.50 Matches: 67
Percent Similarity: 35.94% Conservative: 25
Best Local Similarity: 26.17% Mismatches: 76
Query Match: 12.41% Indels: 88
DB: 1 Gaps: 14

US-10-081-864-14 (1-678) x EXTN_SORBI (1-283)

QY 5 CTTCTCCGAGAGCTCATCAGGTTTCATCGCTTCAAGTGGCATGGAGGACCG 64
Db 82 ProProAlaThrProLysProThr-----Pro 91
QY 65 TGAACGCCACG---AGTTGAGATGAGGCGGAGGCGGAGGCGCCCTACGAGGCG 121
Db 92 ProThrTyrThrProSerProLysProLysSerProValTyrProProProLysAla 111

```

RESULT 9

```

VGLX_HSVB
ID VGLX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
QY 122 ACAACACCGTGAAGTGAAGTGAACCAAGGCGGCCCTGCGCTTCCCTGGACATCC 181
Db 112 SerThrPro-----ProThrTyrThrProSerProLysProProAlaThr 126
QY 182 TGTCTCCCGCCAGTCCAGTACGGCTCCAAAGGTGACGTGAAGCACCCCGCGACATCCCG 241
Db 127 LysProProThrTyrProThr---ProLysProProAlaThrLysProProThrProPro 145
QY 242 ACTACAAAGAACTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAGG 301
Db 146 ValTyrThrProSerProLysProProValThrLys----- 157
QY 302 ACGCGCGCGTGGCGACCGTGACCCAGG---ACTCTCTCCCTGCAGGACGGCTTCATCT 358
Db 158 -----ProProThrProLysProProProValTyrThrProAsn----- 171
QY 359 ACAAGGTGAAGTTTCATCGGCGTGAATTCCTCCCTCCGACG-----GCCCGGTGATGC 409
Db 172 -----ProLysProProValThrLysProProThrHisThrProSerPro----- 186
QY 410 AGAAGAGACCATGGTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGGACCGCGTGC 469
Db 187 -----LysProProThrSerLysProProProValTyrThrProSer----- 201
QY 470 TGAAGGCGGAGATCCACAAAGCGCTGAAGCTGAAGGAGCGCGCCACTACCTGCTGGAGT 529
Db 202 -----ProLysProProLysPro----- 207
QY 530 TCAAGTCCATCATCATG-----CCAAAGAGCGCGTGCAGCTGCCCGGTACTACT 580
Db 208 ---SerProProThrTyrThrProThrProLysProProAlaThrLysProProThrSer 226
QY 581 ACGTGGACA----- 589
Db 227 ThrProThrHisProLysProThrProHisThrProTyrProGlnAlaHisProProThr 246
QY 590 -----CCAAGCTGGACATCACT---CCACACAGGAGACTACACCA 628
Db 247 TyrLysProAlaProLysProSerProProAlaProThrProProThrTyrThrProPro 266
QY 629 TCGTGGACCATAGCGGACCGAGGCGCGCGACCGACCTGTCTCT 674
Db 267 ValSerHisThrProSerSerPro-ProProProProProProPro 281

```


Db 2136 ThrAlaThrThrIleThr-----AlaThrGlySerThrThrAsnProSerSerThr 2152
 QY 305 GCGGCGTGGGACCGTGCAGCCAGGACCTCCCTGCGAGGACGGTCTTCATCATCAAGG 364
 Db 2153 ProGlyThrThrPro-----IleProProValLeuThrThrThrAlaThrThrPro 2169
 QY 365 TGAAGTTCATCGCGTGAACCTCCCTCGAGCGCCCGCGTGATCGCAGAGAAGACCATGG 424
 Db 2170 AlaAlaThrSerThrValThrProSerSerAlaLeuGlyThrThrHisThrPro--- 2188
 QY 425 GCTGGGAGCCCTCACCAGCGCGCTGTACCCCGGAGCGGCTGCTCAAGGGCGAGATCC 484
 Db 2189 -----ProValProAsn-----ThrThrAlaThrThrHisGlyA-gser 2201
 QY 485 ACAAGGCGCTGAAGCTGAAGAGCGCGCCACTACCTGGTGGAGTTCAAGT----- 535
 Db 2202 LeuSerProSerProHisThrValCysThrAlaThrThrSerAlaThrSerGlyIle 2221
 QY 536 -----COATCTACATGCGCAAGAGCGCGCGTGCAGCTGCCCG 571
 Db 2222 LeuGlyThrThrHisIleThrGluProSerThr-----GlyThrSerHisThrPro 2238
 QY 572 GCTACTACTACCTGG-----ACCAAGCTGGGACATCACCT---CCACA 613
 Db 2239 AlaAlaThrThrGlyThrThrGlnHisSerThrProAlaLeuSerProHisProSer 2258
 QY 614 ACGAGGACTACA-----CCATCGTGGAGCAGTACGAGCGCAGCGCGCCACAC 667
 Db 2259 SerArgThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrAla 2278
 QY 668 TGTTCC 673
 Db 2279 ThrSer 2280
 RESULT 11
 SFPO_HUMAN
 ID SFPO_HUMAN STANDARD; PRT; 707 AA.
 DC P23246; P30808; PRT;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
 binding protein-associated splicing factor) (PTB-associated splicing
 factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
 GN SFPO OR PSF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal brain;
 RX MEDLINE=93194059; PubMed=8443401;
 RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
 RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
 factor.";
 RL Genes Dev. 7:393-406(1993).
 RN [2]
 RP SEQUENCE OF 312-707 FROM N.A.
 RC TISSUE=Fetal skeletal muscle;
 RX MEDLINE=90091812; PubMed=2480877;
 RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh P.S.;
 RT "Cloning and characterization of a myoblast cell surface antigen
 defined by 24.1D5 monoclonal antibody.";
 RL Development 105:723-731(1998).
 RN [3]
 RP SEQUENCE OF 48-68 AND 213-246.
 RX MEDLINE=93176127; PubMed=8439294;
 RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
 RT "Purification and characterization of a DNA-binding heterodimer of 52
 and 100 kDa from HeLa cells.";
 RL Biochem. J. 290:267-272(1993).
 CC -!- FUNCTION: Essential pre-mRNA splicing factor required early in

spliceosome formation. Binds to the mammalian polypyrimidine
 tracts. Forms a complex with the polypyrimidine tract-binding
 protein (PTB). Seems to also bind DNA.
 CC -!- SUBUNIT: Heterotrimer of two 52 kDa and two 100 kDa subunits.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=P23246-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P23246-2; Sequence=VSP_005855;
 CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
 CC -!- CAUTION: Was originally (Ref.2) thought to be myoblast cell
 CC surface antigen 24.1D5 and a possible membrane-bound protein
 CC ectokinase.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X70944; CAA50283.1; -;
 CC EMBL; X16850; CAA34747.1; -;
 CC PIR; A46302; A46302.
 CC PIR; S29770; S29770.
 CC HSSP; P11940; 1CVJ.
 CC SWISS-2DPAGE; P23246; HUMAN.
 CC Genew; HGNC:10774; SFPQ.
 CC GK; P23246; -;
 CC MIM; 605199; -;
 CC GO; GO:0008248; P:pre-mRNA splicing factor activity; TAS.
 CC GO; GO:0006397; P:mRNA processing; TAS.
 CC GO; GO:0006371; P:mRNA splicing; TAS.
 CC InterPro; IPR000504; RNA_rec_mot.
 CC Pfam; PF00076; rrm; 2.
 CC SMART; SM00360; RRM; 2.
 CC PROSITE; PS00102; RRM; 2.
 CC PROSITE; PS00030; RRM_RNP_1; 1.
 CC Nuclear protein; RNA-Binding; DNA-binding; mRNA splicing; Repeat;
 KW Alternative splicing.
 FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
 FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
 FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
 FT REPEAT 9 11 1.
 FT REPEAT 19 21 2.
 FT REPEAT 25 27 3.
 FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
 FT DOMAIN 10 15 POLY-GLY.
 FT DOMAIN 20 27 POLY-GLY.
 FT DOMAIN 56 65 POLY-PRO.
 FT DOMAIN 67 71 POLY-GLN.
 FT DOMAIN 95 98 POLY-GLN.
 FT DOMAIN 99 103 POLY-PRO.
 FT DOMAIN 184 188 POLY-PRO.
 FT DOMAIN 571 574 POLY-ARG.
 FT DOMAIN 613 616 POLY-GLY.
 FT DOMAIN 635 641 POLY-GLY.
 FT VARSPLIC 663 707
 FT RTERRFGGAGPVGCGQPRGMGPOTAGYGRGREYEGPNK
 FT KPRF -> VRVIDVG (in isoform Short).
 FT /FTid=VSP_005855.
 FT G -> R (IN REF. 3).
 FT CONFLICT 243 243 6D8D5EA95E235847 CRC64;
 FT SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;
 Alignment Scores:
 Pred. No.: 0.0278
 Score: 150.00
 Percent Similarity: 35.09%
 Best Local Similarity: 29.39%
 Query Match: 12.13%
 Length: 707
 Matches: 67
 Conservative: 13
 Mismatches: 65
 Indels: 84

DB:	1	Gaps:	14
US-10-081-864-14 (1-678) x SFPQ_HUMAN (1-707)			
QY	105	CGGCGCTACAGAGGCGGCAACACACCGTGAAGTTGAAGTGAACAGGCGGCGCCCTGC- 163	
Db	83	ProProHisGlnProProHisProGlnProHisGlnGlnGlnProProProPro 102	
QY	164	-----CCTTCCGCTGGGACA--- 178	
Db	103	ProGlnAspSerSerLysProValValAlaGlnGlyProGlyProAlaProGlyValGly 122	
QY	179	TCTGTCTCCCGCAGTCCAGTACGCTCCAGGTGTACGTGAAGCACCCTCCGACATCC 238	
Db	123	SerAlaProProAlaSerSerAla-----ProProAlaThr 135	
QY	239	CGGACTACAGAAGCTGCTCTCCCGAGGCTTCAAGTGGAGCGCGTGAATCTCG 298	
Db	136	ProProThrSerGlyAlaPro--ProGlySerGlyProGly----- 148	
QY	299	AGGACGGCGGCTGGCGGCGGCGGCTCCCTCCCTGAGGAGCGCTGCT----- 352	
Db	149	-----ProThrProThrProProProAlaValThrSerAlaProPro 162	
QY	353	-----TCATCTACAAGTGAAGTTCATCGGCGTGAATCTCCCTCCGACGCGCCCG 403	
Db	163	GlyAlaProProThrProProSerSerGlyValProThrProProGlnAla--- 181	
QY	404	TGATCAGAGAAGACCATGGCGTGGAGCGCTCCACGAGCGCTGTACCCCGCGAGC 463	
Db	182	-----GlyGlyProProProProProAlaValProGlyPro 194	
QY	464	GCCTGCT-----GAAGGCGGATCCACAGGCGCTGAAGCTGAAGGCGGCGGCG 514	
Db	195	Gly-ProGlyProGlyGlnGlyProGlyProGlyProGlyGly---GlyLysMetPr 213	
QY	515	A-----CTACTGTGGGA-----GTTCA 532	
Db	213	OGlyGlyProLysProGlyGlyProGlyLeuSerThrProGlyGlyHisProLysPr 233	
QY	533	AGTCCATCTACATGGCGAGACCGCGTGCAGCTGCCGGCTACTACTACTGACACCA 592	
Db	233	opzHisargGlyGlyGlyGluProArgly-----GlyArgG 246	
QY	593	AGCTGACATCACTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGAGCGCACCG 652	
Db	246	n-----HisHisProProTyHisGlnGlnHis-----HisG 257	
QY	653	AGGCGCGCCACCACTGTTCCT 674	
Db	257	nGlyProProProGlyGlyPro 264	
RESULT 12			
IF2_STRAW		STANDARD;	PRT; 1046 AA.
ID	IF2_STRAW		
AC	Q82K53;		
DT	15-MAR-2004 (Rel. 43, Created)		
DT	15-MAR-2004 (Rel. 43, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Translation initiation factor IF-2.		
GN	INFB OR SAV2551.		
OS	Streptomyces avermitilis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=33903;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=21477403; PubMed=11572948;		
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,		
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,		
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.,		
RT	"Genome sequence of an industrial microorganism Streptomyces		

RT	avermitilis: deducing the ability of producing secondary metabolites."		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).		
RN	(2)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
EX	MEDLINE=22608306; PubMed=12692562;		
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,		
RA	Sakaki Y., Hattori M., Omura S.;		
RT	"Complete genome sequences and comparative analysis of the industrial microorganism Streptomyces avermitilis."		
RL	Nat. Biotechnol. 21:526-531 (2003).		
CC	!- FUNCTION: One of the essential components for the initiation of protein synthesis. Protects formylmethionyl-tRNA from spontaneous hydrolysis and promotes its binding to the 30S ribosomal subunits.		
CC	Also involved in the hydrolysis of GTP during the formation of the 70S ribosomal complex (By similarity).		
CC	70S ribosomal complex (By similarity).		
CC	!- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	!- SIMILARITY: Belongs to the IF-2 family.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AP005031; BAC70262.1; -.		
DR	HAWAP; MF_00100; -; 1.		
DR	InterPro; IPR001851; Bac_inmem_transp.		
DR	InterPro; IPR004161; EFTU_D2.		
DR	InterPro; IPR000795; EF_GTPbind.		
DR	InterPro; IPR003577; GTPase_Ras.		
DR	InterPro; IPR000178; IF2_N.		
DR	InterPro; IPR008847; IF2_N.		
DR	InterPro; IPR002965; P-rich_extensn.		
DR	InterPro; IPR001806; Ras_trnsfrmg.		
DR	InterPro; IPR005225; Small_GTP.		
DR	InterPro; IPR009000; Translat_factor.		
DR	Pfam; PF00009; GTP_EFTU; 1.		
DR	Pfam; PF03144; GTP_EFTU_D2; 1.		
DR	Pfam; PF04760; IF2_N; 2.		
DR	PRINTS; PR00315; ELONGATNFCT.		
DR	PRINTS; PR01217; PRICHEXTENS.		
DR	PRINTS; PR00449; RASTRNSFRMG.		
DR	ProDom; PD186100; IF2; 1.		
DR	SMART; SM00173; Ras; 1.		
DR	TIGRFAMs; TIGR00487; IF-2; 1.		
DR	TIGRFAMs; TIGR00231; small_GTP; 1.		
DR	PROSITE; PS01176; IF2; FALSE_NEG.		
KW	Initiation factor; Protein biosynthesis; GTP-binding;		
KW	Complete proteome.		
FT	DOMAIN 542 694		
FT	NP_BIND 548 555		
FT	NP_BIND 598 602		
FT	NP_BIND 652 655		
SQ	SEQUENCE 1046 AA; 106882 MW; 62B12E2DFA3AA595 CRC64;		
Alignment Scores:			
Pred. No.:	0.0284	Length:	1046
Score:	150.00	Matches:	69
Percent Similarity:	38.64%	Conservative:	16
Best Local Similarity:	31.36%	Mismatches:	91
Query Match:	12.00%	Indels:	44
DB:	1	Gaps:	12
US-10-081-864-14 (1-678) x IF2_STRAW (1-1046)			
QY	655	CCTCGGTCGGCTCGTACTGCTCCAGCAGTGTAGTCTCGTGTGGAGGTGATGCCA 596	
Db	101	ProAlaAlaAlaProAlaAlaPro-----GlyProArgProGly-----Pro 114	

Db 1392 SerProThrThrSerThrProSerThrProGlnProThr-----SerSer 1408
QY 227 CCGCGGACATCCCGGACTACAGAAGCTGTCCTCCCGAGGGCTCAAGTGGGAGCGG 286
Db 1409 ProThrThrLeuProThrThr-----SerProLeuThrSerSerAlaThrSer 1424
QY 287 TGATGAACTTCGAGGACGCGCGCTGGCGACCGTGCAGGACCTCCCTGCGAGGACG 346
Db 1425 -----ProThrThrSerHisIleThrSer 1432
QY 347 GCTGCTTCATCTACAGGTGAAGTTCATCGGCGTGAATCTCCCGAGCGCCCGGTGA 406
Db 1433 ThrValSerProThrThrSerProThrThrSerProThrThrSerPro--- 1451
QY 407 TCGAGAGAAGACCATGGCTGGAGGCTCCACGAGCGCTGTACCCCGCGAGCGG 466
Db 1452 -----ThrThrSerThr--- 1455
QY 467 TCGTGAAGGCGGAGATCCACAGGCGCTGAAGCTGAAGGACGCGCGGCTACCTCGGTGG 526
Db 1456 -----ThrSerPro-----ThrThrSerThrThrSerPro 1465
QY 527 AGTTCAGTCCATCTACATGCGGACGAGAGCGGCTGCAGTGCCTGCTACTACTAGGTGG 586
Db 1466 ThrProSerProThrThrSerThrThrSerProThrProSerProThrThrSerThr 1485
QY 587 ACACCAAGCTGGACATCCACTCCACAGGAGGACTACACCATCGTGGAGCAGTACGAGC 646
Db 1486 SerProThr---ProSerProThrThrSerThrThrSerProThrThrSerProIleThr 1504
QY 647 GCACCGAGGCGCCACCCACC 667
Db 1505 SerProThrThrSerThrThr 1511
RESULT 14
SFRG_HUMAN STANDARD; PRT; 659 AA.
AC Q8N2M8; O96026; Q96DX2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Splicing factor, arginine/serine-rich 16 (Suppressor of white-apricot
DE homolog 2).
GN SFRS16 OR SWAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Yoshiura K., Murray J.C.;
RT "A transcriptional map in the region of 19q13 derived using direct
RT sequencing and exon trapping";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Embryo;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuko Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein V.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
Rana S.S., Joquellano N.A., Peters G.J., Abranson R.D., Mullaly S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
PL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Probably functions as an alternative splicing regulator.
CC May regulate the mRNA splicing of genes such as CLK1. May act by
CC regulating members of the CLK kinase family (By similarity).
CC -!- SUBUNIT: Probably interacts with CLK4 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8N2M8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8N2M8-2; Sequence=VSP_008207, VSP_008208, VSP_008209,
CC VSP_008210;
CC Note=No experimental confirmation available;
CC -!- PTM: Phosphorylated in vitro by CLK4 (By similarity).
CC -!- SIMILARITY: Belongs to the splicing factor SR family.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to some
CC missing sequence in an exon that changes the coding frame.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF042800; AAC82339.1; -;
DR EMBL; AF042810; AAC82340.1; -;
DR EMBL; AF042802; AAC82340.1; JOINED.
DR EMBL; AF042803; AAC82340.1; JOINED.
DR EMBL; AF042804; AAC82340.1; JOINED.
DR EMBL; AF042805; AAC82340.1; JOINED.
DR EMBL; AF042806; AAC82340.1; JOINED.
DR EMBL; AF042807; AAC82340.1; JOINED.
DR EMBL; AF042808; AAC82340.1; JOINED.
DR EMBL; AF042809; AAC82340.1; JOINED.
DR EMBL; AK074590; BAC11078.1; -;
DR EMBL; BC013178; AAH13178.1; -;
DR Genew; HGNC:17731; SFRS16.
KW mRNA processing; mRNA splicing; Nuclear protein; Alternative splicing;
KW Polymorphism.
FT DOMAIN 354 658 ARG-RICH.
FT DOMAIN 362 520 SER-RICH.
FT VARSPLIC 1 334 Missing (in isoform 2).
FT VARSPLIC 335 359 /FTId=VSP_008207.
FT VARSPLIC 456 493 RPALTGPPCPACS (in isoform 2).
FT VARSPLIC 493 SPSHGDEYRGGLRHSSRSRSSLSPSRSL ->
FT (in isoform 2).
FT /FTId=VSP_008208.
FT Missing (in isoform 2).
FT VARSPLIC 494 659 /FTId=VSP_008210.
FT VARSPLIC 159 159 E -> K (in dbSNP:4803794).
FT VARIANT 410 410 S -> P (IN REF. 2).
FT CONFLICT 659 AA; 75214 MW; 1E455FA232BF30D4 CRC64;
SQ SEQUENCE

Alignment Scores:

Pred. No.: 0.0369 Length: 659
 Score: 148.00 Matches: 71
 Percent Similarity: 39.81% Conservative: 15
 Best Local Similarity: 32.87% Mismatches: 74
 Query Match: 11.84% Indels: 56
 DB: 1 Gaps: 11

US-10-081-864-14 (1-678) x SFRG_HUMAN (1-659)

```

QY 577 AGTAGCGGGGAGTTCAGCGGCTTCTTGGCCATGTAGATGCACTTGAATCCACACGAGT 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 AlalaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 343
QY 517 AGTGGCCGCCCTCTTCTGAGTTCAGCGGCTTGT---GGATCTCGCCCTTACGACGCGGT 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 344 Pro-----ProGlnProGlyGlyProAlaProGlyArgAsnAlaSerAlaArgArg 360
QY 460 CGCGGGGTACAGCGGCTCGGTGAGGCGCTCCAGCCCGATGCTTCTTCTGATCAGCG 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 ArgSerSer-----SerSerSerSerSerSer 370
QY 400 GGCGGTGGAGGGAAGTTCAGCGGATGAATTCACCTTGTAGATGAACGACGCGCTCT 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 AlaserArgThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 390
QY 340 GCAGGGAGGTCTGGTCTACGCTCGCCAGCGCGCTCTCGAAGTTCATCAGCGCT 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 GlyGlyTyrTyrArgSerGlyArgHisAlaArgSerArgSerArgSerTrpSerArg 410
QY 280 CCCACTTGAAGCCCTCGGGAAGACACACTTCTTGTAGTCGGGATGTCGGCGGGTGCT 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 ArgSerArgSer---ArgArgTyrSerArgSerArgSerArgGlyArgHisSerGly 429
QY 220 -----TCAGGTACACTTGGAGCGGTACTGGAATCGG 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 GlyGlySerArgAspGlyHisArgTyrSerArgSerProAlaArgArgGlyGlyTrGly 449
QY 187 GGGCAGGATGCCAGCGGAAGGCA----- 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 -----ProArgArgArgSerArgSerArgSerHisSerGlyAspArgTyrArg 465
QY 160 ---GGGGCGCG-----CCTTGG----- 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 ArgGlyArgGlyLeuArgHisSerSerSerArgSerArgSerTrpSerLeu 485
QY 145 TCACCTTCAACTTCACGCTGTGTGGCCCTCGTAGGGCG---GGCCCTCGCCT---CGC 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 SerProSerArgSerArgSerLeuThrArgSerArgSerHisSerProSerProSerGln 505
QY 91 CCTCGATCTCGAACTCGTGGCCGTTTCAGGTGCGCTTCATGCGCACCT 44
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 SerArgSerArgSerArgSerArgSerGlnSerProSerProSerProSer 521

```

RESULT 15

ID MUSA_HUMAN STANDARD; PRT; 1233 AA.
 AC P98088; Q76065; Q13792; Q8WWQ3;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Mucin 5AC (Mucin 5 subtype AC, tracheobronchial) (Tracheobronchial
 DE mucin) (TEM) (Major airway glycoprotein) (Fragment).
 GN MUC5AC (MUC5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-372 FROM N.A.
 RX MEDLINE=21426417; PubMed=11535137;
 RA Escande F., Aubert J.-P., Porchet N., Buisine M.P.;

RT RT Human mucin gene MUC5AC: organization of its 5'-region and central
 RL RL repetitive region.";
 RN RN Biochem. J. 358:763-772(2001).
 RP RP SEQUENCE OF 193-1233 FROM N.A., AND PARTIAL SEQUENCE.
 RX RX TISSUE=lung;
 RA RA MEDLINE=94230376; PubMed=7513696;
 RA Meerman D., Charles P., Daskal E., Polymeropoulos M.H.,
 RA Martin B.M., Rose M.C.;
 RT RT "Cloning and analysis of cDNA encoding a major airway glycoprotein,
 RT human tracheobronchial mucin (MUC5).";
 RL J. Biol. Chem. 269:12932-12939(1994).
 RN RN [3]
 RP RP SEQUENCE OF 153-1233 FROM N.A.
 RX RX TISSUE=Placenta, and Trachea;
 RA RA MEDLINE=98285541; PubMed=9620876;
 RA Buisine M.P., Desseyn J.L., Porchet N., Degand P., Laine A.,
 RA Aubert J.-P.;
 RT RT "Genomic organization of the 3'-region of the human MUC5AC mucin gene:
 RT additional evidence for a common ancestral gene for the 11p15.5 mucin
 RT gene family.";
 RL Biochem. J. 332:729-738(1998).
 RN RN [4]
 RP RP SEQUENCE OF 284-1233 FROM N.A.
 RX RX MEDLINE=95293957; PubMed=7775418;
 RA Lesuffleur T., Roche F., Hill A.S., Lacasa M., Fox M., Swallow D.M.,
 RA Zweibaum A., Real F.X.;
 RT RT "Characterisation of a mucin cDNA clone isolated from HT-29 mucus
 RT secreting cells: The 3' end of MUC5AC?";
 RL J. Biol. Chem. 270:13665-13673(1995).
 CC CC -!- FUNCTION: MAJOR AIRWAY GLYCOPROTEIN.
 CC CC -!- SUBUNIT: Multimeric.
 CC CC -!- TISSUE SPECIFICITY: Tracheobronchial MUCOSAE (NOT SOLELY).
 CC CC -!- SIMILARITY: TO VARIOUS MUCINS.
 CC CC -!- SIMILARITY: Contains 1 WFCD domain.
 CC CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ298319; CAC83676.1; --
 CC EMBL; U06711; AA18431.1; --
 CC EMBL; AJ001402; CAA04737.1; --
 CC EMBL; AJ001403; CAA04738.1; --
 CC EMBL; Z48314; CAA88307.1; --
 CC Genew; HGNC:7515; MUC5AC.
 CC GO; GO:0005201; P:extracellular matrix structural constituent; TAS.
 CC GO; GO:007155; P:cell adhesion; NAS.
 CC InterPro; IPR006208; Cys_knot.
 CC InterPro; IPR006207; Cys_knot_C.
 CC InterPro; IPR002919; TIL_Cysrich.
 CC InterPro; IPR001007; VWF_C.
 CC InterPro; IPR001846; VWF_D.
 CC Pfam; PF00007; Cys_Knot_1.
 CC Pfam; PF00093; vwd; 1.
 CC Pfam; PF00094; vwd; 1.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS01208; VWF_C_1; 2.
 CC PROSITE; PS01084; VWF_C_2; 2.
 CC Repeat; Glycoprotein.
 CC NON TER 1
 CC DOMAIN 431 496
 CC DOMAIN 500 659
 CC DOMAIN 688 712
 CC REPEAT 688 692
 CC 5 X 5 AA TANDEM REPEATS OF T-T-V-G-P/S.
 CC 1.

344	Qy	ACGGCTGCTTCATCTACAAGGTGAAGTTCATCGCGCTGAACCTCCCTCCGAGCGCCCGC	403
80	Db	SerThrAlaSerThrThr-----SerGlyProGlyThrThrPro	92
404	Qy	TGATGCAGAGAAGACCATGGCTGGAGAGCCCTCCACCGAGCGCTGTACCCCGCGACG	463
93	Db	-----SerProValProThrThrSerThrThrSer-----	102
464	Qy	GCCTGCTGAAGGCGGAGATCCACAAGGCCCTGAAGCTGAAGGAGCGGGCCACACTACTGG	523
103	Db	-----AlaProThrThrSerThrThrSerAlaSerThrAlaSerThr-----	117
524	Qy	TGGAGTTCAGTCCATCTACATGGCCCAAGAGCCGCTGCAGCTCCCGGCTACTACTACG	583
118	Db	-----SerGlyProGlyThr-----SerLeuSerProValProThrThrSerThr	132
584	Qy	TGGACACCAAGCTGGACATCACCTCCCAACAGAGGACTACA	625
133	Db	-----ThrSerAlaProThrThrSerThrThrSerGlyProGlyThrThr	147
626	Qy	-----CCATCGTGGACAGTACGAGCGCACCGAGGGCGGCCACCCACC	667
148	Db	ProSerProValProThrThrSerThrThrSerAlaProThrThrSerThr	165

Search completed: July 29, 2004, 14:33:24
Job time : 20.7906 secs

[illegible]

Blank sheet USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:27:25 ; Search time 40.3863 Seconds
(without alignments)
10593.763 Million call updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggctctctcgagaagt.....gcaccacctgttctgttaa 678

Scoring table: BLOSUM62 Xgapop 10.0 Xgapext 0.5
Ygapop 10.0 Ygapext 0.5
Fgapop 6.0 Fgapext 7.0
Delop 6.0 Delext 7.0

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10081864/runat_29072004_150506_25475/app_query.fasta_1.1742
-DB=SPREMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 -CGN_1_1_122 @runat_29072004_150506_25475 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mbc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	1191	96.3	225 5 Q9U6Y8	Q9U6Y8 discosoma s

ALIGNMENTS

RESULT 1		PRELIMINARY:		PRT:		225 AA.	
Q9U6Y8	Q9U6Y8	ID	Q9U6Y8				
AC	Q9U6Y8						
DT	01-MAY-2000						
DT	01-MAY-2000						
DT	01-OCT-2003						
DE	Fluorescent protein FP583.						
OS	Discosoma sp.						
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;						
OC	Discosomatidae; Discosoma.						
OX	NCBI_TaxID=86600;						
RP	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=99436614; PubMed=10504696;						
RA	Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,						
RA	Markelov M.L., Lukyanov S.A.;						
RT	"Fluorescent proteins from nonbioluminescent Anthozoa species.";						
RL	Nat. Biotechnol. 17:969-973(1999).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RA	Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,						
RA	Markelov M.L., Lukyanov S.A.;						
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AF169419; AAF03369.1; -						

Q9gti7 discosoma s
Q9sp04 gonipora t
Q9u6y7 discosoma s
Q963f5 montastraea
Q95ua7 montastraea
Q7z0w4 montastraea
Q7z0w6 montastraea
Q962p9 montastraea
Q7z0w8 montastraea
Q7z0w5 montastraea
Q7z0w7 montastraea
Q8t5f2 montastraea
Q8t6u0 dendronephth
Q9u6y3 ciavularia
Q95vt0 montastraea
Q8mu47 montastraea
Q7z0w9 montastraea
Q8mu48 montastraea
Q8t5f1 montastraea
Q8t5f8 parasicyoni
Q8t6j8 trachyphyl
Q8mma2 agaricia fr
Q8t5f0 scolymia cu
Q8mmal agaricia ag
Q9gpi6 anemonia su
Q8t6t8 discosoma s
Q8t5e9 ricordea fl
Q8t5e8 ricordea fl
Q8mu46 ricordea fl
Q95w11 condylactis
Q8t5f3 scolymia cu
Q8t5f4 radianthus
Q8mu45 condylactis
Q8t6t9 radianthus
Q8t5e7 condylactis
Q9u6y6 anemonia ma
Q8t6l8 meandrina m
Q8t6l7 meandrina m
Q9u6y5 zoanthus sp
Q9u6y4 zoanthus sp

Db 181 TyMetValLysLysProSerValGlnLeuProGlyTyrTyrValAspSerLysLeu 200
Qy 598 GACATCACCTCCACACAGGAGTACTACCATCGTGGAGCAGTACGAGCCAGGAGGC 657
Db 201 AspMetThrSerHisAsnGluAspTyrThrValValGluGlnTyrGluLysThrGlnGly 220
Qy 658 CGCCACCACTGTCCTG 675
Db 221 ArgHisHisProPheille 226
RESULT 3
Q95P04 PRELIMINARY; PRT; 221 AA.
AC Q95P04; 201-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Gonipora tenuidens.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Fungiina; Poritidae; Gonipora.
OX NCBI_TaxID=75301;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Fradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF383156; AAL27542.1; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 24918 MW; 93F9F4B5C2003CB4 CRC64;

Alignment Scores:
Pred. No.: 7,37e-51 Length: 221
Score: 795.00 Matches: 144
Percent Similarity: 80.48% Conservative: 25
Best Local Similarity: 68.57% Mismatches: 41
Query Match: 64.27% Indels: 0
DB: 5 Gaps: 0
US-10-081-864-14 (1-678) x Q95P04 (1-221)
Qy 16 AACGTATACACGAGTTCATCGCTTCAGGTCCGATGAGGAGGACCGTGAACGCCAC 75
Db 2 SerValIleAlaLysGlnMetThrTyrLysValTyrMetSerGlyThrValAsnGlyHis 21
Qy 76 GAGTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGCGCACACACCGTGAAG 135
Db 22 TyrPheGluValGlnGlyAspGlyLysGlyLysProTyrGluGluGlnThrValLys 41
Qy 136 TTGAAGGTGACCAAGGCGCGCCCTCGCTTCGCTGGACATCTGTCGCCCGCCAGTTC 195
Db 42 LeuThrValThrLysGlyGlyProLeuProPheAlaTyrAspIleLeuSerProGlnSer 61
Qy 196 CAGTACGCTCCAAAGTGTACGTGACGACCCCGCCGATCCCGACTACAGAGCTG 255
Db 62 GlnTyrGlySerIleProPheThrLysTyrProGluAspIleProAspTyrValLysGln 81
Qy 256 TCCTTCCCGGAGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGGACGCGCGGTGGG 315
Db 82 SerPheProGluGlyTyrThrTyrGluArgIleMetAsnPheGluAspGlyAlaValCys 101
Qy 316 ACCTGACCCAGACTCTCTCCGAGGAGCGGTGCTTCACTCAAGGTGAAGTTCATC 375
Db 102 ThrValSerAsnAspSerIleGlnGlyAsnCysPheIleTyrAsnValLysPheSer 121

Qy 376 GCGGTGACCTTCCCTCCGAGCGCCCGTGTATGACAGAGAGACCATGGCTGGAGGCC 435
Db 122 GlyLeuAsnPheProProAsnGlyProValMetGlnLysLysThrGlnGlyTyrGluPro 141
Qy 436 TCCACCGAGCGCTGTACCCCGGAGCGCGTGTGAGGCGGAGATCCACAGGCGCTG 495
Db 142 AsnThrGluArgLeuPheAlaArgAspGlyMetLeuIleGlyAsnAsnPheMetAlaLeu 161
Qy 496 AAGCTGAAGGACGCGCGGCTACTACTACTGACACCAAGCTGGACATCACCTCCCAAC 555
Db 162 LysLeuGluGlyGlyGlyHisTyrLeuCysGluPheLysSerThrTyrLysAlaLysLys 181
Qy 556 CCGGTGACGCTCCCGGCTACTACTACTGACACCAAGCTGGACATCACCTCCCAAC 615
Db 182 ProValLysMetProGlyTyrHisTyrValAspArgLysLeuAspValThrAsnHisAsn 201
Qy 616 GAGGACTACACATCGTGGAGCAGTACGAG 645
Db 202 IleAspTyrThrSerValGluGlnCysGlu 211
RESULT 4
Q9U6Y7 PRELIMINARY; PRT; 232 AA.
AC Q9U6Y7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP483.
OS Discosoma striata.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=105400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
Markelov M.L., Lukyanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
DR EMBL; AF168420; AAF03370.1; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 232 AA; 26435 MW; AA8F18EE283CE4D CRC64;

Alignment Scores:
Pred. No.: 9.15e-46 Length: 232
Score: 726.50 Matches: 132
Percent Similarity: 76.00% Conservative: 39
Best Local Similarity: 58.67% Mismatches: 53
Query Match: 58.73% Indels: 1
DB: 5 Gaps: 1
US-10-081-864-14 (1-678) x Q9U6Y7 (1-232)

Qy 1 ATGGCTCTCCGAGAGCTCATCCAGTTCATCGCTTCAAGGTGCGCATGAGGCG 60
Db 1 MetSerCysSerLysSerValIleLysGluGluMetLeuLeuAspLeuHisLeuGluGly 20
Qy 61 ACCGTGAACGCGCAGGTTGAGATCGAGGCGAGGCGGAGGCGGCGGCGGCGGCG 120
Db 21 ThrPheAsnGlyHisTyrPheGluIleLysGlyLysGlyGlnProAsnGluGly 40
Qy 121 CACACACCGTGAAGTTCAGGTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCG 180
Db 41 ThrAsnThrValThrLeuGluValThrLysGlyGlyProLeuProPheGlyTyrHisIle 60
Qy 181 CTGTCCCCCGAGTTCAGTACGCTCCAGGTGACGTGAAGACCCCGCGCATCCCC 240
Db 61 LeuCysProGlnPheGlnTyrGlyAsnLysAlaPheValHisHisProAspAsnIleHis 80

241 GACTACAAAGAGCTGCTCTCCCGAGCGGCTTCAAGTGGAGCGCGGTGATGAATCTCGAG 300
 81 AspyrLeuLysLeuSerPheProGluGlyTyrThrTrpGluArgSerMetHisPheGlu 100
 301 GACGGCGGCGGCGAGCGGTGACCCAGCACTCTCCCTCCGACGCGCGGTGATGCAAGAAC 360
 101 AspGlyGlyLeuCysCysIleThrAsnAspIleSerLeuThrGlyAsnCysPheTyrTyr 120
 361 AAGTGAAGTTCATCGCGTGAATCTCCCTCCGACGCGCGCGGTGATGCAAGAAC 420
 121 AspileLysPheThrGlyLeuAsnAspPheProAsnGlyProValValGlnLysLysThr 140
 421 ATGGCGTGGAGGCTCCACGAGCGCTGTACCCCGCGACGCGGTGCTGAAGGCGGAG 480
 141 ThrGlyTrpGluProSerThrGluArgLeuTyrProArgAspGlyValLeuIleGlyAsp 160
 481 ATCCACAAAGCGCTGAAGTGAAGGAGCGCGCGCACTACTCTGTTGGAGTTCAGATCCATC 540
 161 IleHisIleAlaLeuThrValGluGlyGlyHisTyrAlaCysAspIleLysThrVal 180
 541 TACATGCCCAAGAG---CCGCTGACGCTCCCGGCTACTACTACGTGGAGCACCAAGCTG 597
 181 TyrArgAlaLysLysAlaLeuLysMetProGlyTyrHisTyrValAspThrLysLeu 200
 598 GACATCACTCCCAACAGGAGTACACCATCTGTGGAGCAGTACGACGCGACCGAGGGC 657
 201 ValIleIlePheAsnAspLysGluPheMetLysValGluGluHisGluIleAlaValAla 220
 658 CCGCACCACTGTTTC 672
 221 ArgHisProPhe 225

RESULT 5
 Q963F5 PRELIMINARY; PRT; 225 AA.
 ID Q963F5
 AC Q963F5
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Green fluorescent protein.
 OS Montastraea cavernosa (great star coral).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
 OC Faviina; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
 RA Falkowski P., Gorbunov M., Kolber Z.;
 RI "Green fluorescent proteins in Caribbean Scleractinian corals.";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF394683; AAK62982.2;
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP.1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 DR NON_TER 225
 SQ SEQUENCE 225 AA; 25847 MW; 77DE7D7C616929AF CRC64;

Alignment Scores:
 Pred. No.: 1,31e-42 Length: 225
 Score: 684.00 Matches: 119
 Percent Similarity: 74.77% Conservative: 44
 Best Local Similarity: 54.59% Mismatches: 55
 Query Match: 55.30% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x Q963F5 (1-225)

76 GAGTTCGAGATCGAGGCGGAGCGCGCCCTACGAGGCGCACACACCGCTGAAG 135
 22 LysPheValIleLysGlyGluGlyGlyProGluGlyThrGlnThrIleAsn 41
 136 ITGAAGGTGACACGAGGCGCGCCCTGCTCCCTTCGCTGGGACATCTGTCCCCCAGTTC 195
 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrSerAlaPhe 61
 196 CAGTACGCTCAAGGTGTACGTGAAGCACCCCGCGACATCCCGCTCAACAAGCTG 255
 62 GlnTyrGlyAsnArgValPheThrLysTyrProAspAspIleProAspTyrPheLysGln 81
 256 TCCTTCCCGGAGGCTTCAAGTGGGAGCGCTGTAGTAACTTCGAGGACGCGCGGTGGCG 315
 82 ThrPheProGluGlyTyrSerTyrGluArgIleMetAlaTyrGluAspGlnSerIleCys 101
 316 ACCGTGACCCAGGACTCTCCCTCGAGGACCGCTGCTTCATCTACAAGGTGAAGTTCATC 375
 102 ThrAlaThrSerAspIleLysMetGluGlyAspCysPheIleTyrGluIleGlnPheHis 121
 376 GCGTGAAGTTCCTCCCTCGAGCGCGCTGTAGTGAAGAACAGACCATCGGCTGGAGGCC 435
 122 GlyValAsnPheProProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
 436 TCCACCGAGCGCTGTACCCCGCGAGCGCTGTGTGAAGGCGGAGATCCACAAGGCCCTG 495
 142 SerThrGluLysMetTyrValArgAspGlyValLeuLysGlyAspValAsnMetAlaLeu 161
 496 AAGCTGAAGGACGCGCGCTACTACTGTGTGAGTTCAGTCCATCTACATCGCCCAAGAG 555
 162 LeuLeuGluGlyGlyHisTyrArgCysAspPheArgSerThrTyrLysAlaLysLys 181
 556 CCCGTGACGCTCCCGGCTACTACTGTGTGAGTTCAGTCCATCTACATCGCCCAAGAG 615
 182 ArgValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeuSerHisAsp 201
 616 GAGGACTACCATCGTGGAGCAGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 669
 202 AsnAspTyrAsnThrValLysLeuSerGluAspAlaGluAlaArgTyrSerMet 219

RESULT 6
 Q95UA7 PRELIMINARY; PRT; 225 AA.
 ID Q95UA7
 AC Q95UA7
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Cyan fluorescent protein (fragment).
 OS Montastraea cavernosa (great star coral).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
 OC Faviina; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Falkowski P.G., Sun Y.;
 RI "Montastraea cavernosa fluorescent protein.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY056460; AAL17905.1;
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP.1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 DR NON_TER 225
 SQ SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Alignment Scores:
 Pred. No.: 6.1e-42 Length: 225
 Score: 675.00 Matches: 122
 Percent Similarity: 73.49% Conservative: 36
 Best Local Similarity: 56.74% Mismatches: 57

Db 202 SerHisAspLysAspTyrAsnLysValLysLeuTyrGluHisAlaGluAlaHisSerGly 221
QY 667 CTG 669
Db 222 Leu 222
RESULT 11
Q7ZOW5 PRELIMINARY; PRT; 225 AA.
AC Q7ZOW5; 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein.
OS Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc5;
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V.; Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great
Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181556; AAO61602.1; -.
SQ SEQUENCE 225 AA; 25843 MW; 13708587B7D93E35 CRC64;

Alignment Scores:
Pred. No.: 1.02e-41 Length: 225
Score: 672.00 Matches: 121
Percent Similarity: 73.02% Conservative: 36
Best Local Similarity: 56.28% Mismatches: 58
Query Match: 54.32% Indels: 0
DB: Gaps: 0

US-10-081-864-14 (1-678) x Q7ZOW5 (1-225)

QY 16 AACGTCAATCAACCGAGTTCATCGCTTCAAGTGGCATGAGGGCACCGTGAACGGCCAC 75
Db 2 SerValIleLysSerValMetLysIleLysLeuHisMetAspGlyIleValAsnGlyHis 21
QY 76 GAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGGCCACACACCGTGAAG 135
Db 22 LysPheMetIleThrGlyGluGlyLysProPheGluGlyThrHisThrIle 41
QY 136 TTGAAGGTGACCAAGGGCGCGCCCTGCTTGGCTGGGACATCCTGTCCTCCCGAGTTC 195
Db 42 LeuLysValLysGluGlyGlyProLeuProPheAlaTyrAspIleLeuThrThrAlaPhe 61
QY 196 CAGTACGGCTCCAAAGTGTACGTGAAGCACCCCGCGGACATCCCGACTACAGAGCTG 255
Db 62 GlnTyrGlyAsnArgValPheThrLysTyrProLysAspIleProAspTyrPheLysGln 81
QY 256 TCCTTCCCGAGGGCTTCAAGTGGCGAGCGCGTGAATCTTCGAGGACGGCGCGTGGCG 315
Db 82 SerPheProGluGlyTyrSerTrpGluArgSerMetThrPheGluAspGlnGlyValCys 101
QY 316 ACCGTGACCCAGGACTCCTCCTCGAGGACGGCTCTCATCTCAAGTGAAGTTCATC 375
Db 102 ThrValThrSerAspIleLysLeuGlyAspCysPhePheTyrGluIleArgPheTyr 121
QY 376 GCGGTGAATTCCTCCCGACCGCGCGCTGATGCGAGAGAACCATGGCTGGGAGGCC 435
Db 122 GlyValAsnPheProSerSerGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY 436 TCACCGAGCGCTGTACCCCGCGGCGCTGCTGAGGGCGAGATCCACAGGCCCTG 495
Db 142 SerThrGluAsnMetTyrValArgAspGlyValLeuLeuGlyAspValSerArgThrLeu 161
QY 496 AAGCTGAAGGACGGCGCGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGCGCCAGAG 555

Q7ZOW8 PRELIMINARY; PRT; 227 AA.
AC Q7ZOW8; 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc2;
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V.; Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great
Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181553; AAO61599.1; -.
SQ SEQUENCE 227 AA; 26017 MW; 5E312C54EA47F589 CRC64;

Alignment Scores:
Pred. No.: 9.38e-42 Length: 227
Score: 672.50 Matches: 120
Percent Similarity: 74.21% Conservative: 44
Best Local Similarity: 54.30% Mismatches: 54
Query Match: 54.37% Indels: 3
DB: Gaps: 1

US-10-081-864-14 (1-678) x Q7ZOW8 (1-227)

QY 16 AACGTCAATCAACCGAGTTCATCGCTTCAAGTGGCATGAGGGCACCGTGAACGGCCAC 75
Db 2 SerValIleLysProAspMetLysIleLysLeuArgMetGluGlyAlaValAsnGlyHis 21
QY 76 GAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGGCCACACACCGTGAAG 135
Db 22 LysPheValIleGluGlyAspGlyLysGlyLysProPheGluGlyThrGlnSerMetAsp 41
QY 136 TTGAAGGTGACCAAGGGCGCGCCCTGCTTGGCTGGGACATCCTGTCCTCCCGAGTTC 195
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrThrValPhe 61
QY 196 CAGTACGGCTCCAAAGTGTACGTGAAGCACCCCGCGGACATCCCGACTACAGAGCTG 255
Db 62 AspTyrGlyAsnArgValPheAlaLysTyrProGlnAspIleProAspTyrPheLysGln 81
QY 256 TCCTTCCCGAGGGCTTCAAGTGGCGAGCGCGTGAATCTTCGAGGACGGCGCGTGGCG 315
Db 82 ThrPheProGluGlyTyrSerTrpGluArgSerMetThrTyrGluAspGlnGlyIleCys 101
QY 316 ACCGTGACCCAGGACTCCTCCCTGCGAG- - - - -GACGGCTGCTTCATCTCAAGGTG 366
Db 102 ValAlaThrAsnAspIleThrLeuMetLysGlyValAspAspCysPheValTyrLysIle 121
QY 367 AAGTTCATCGCGTGAATCTCCCTCCGACGGCGCGCTGATGCGAGAGAACCATGGCG 426
Db 122 ArgPheAspGlyValAsnPheProAlaAsnGlyProValMetGlnLysLysThrLeuLys 141
QY 427 TGGAGGCGCTCCACCGAGCGCTGTACCCCGCGGACGGCTGCTGAGGGCGAGATCCAC 486
Db 142 TrpGluProSerThrGluLysMetTyrValArgAspGlyValLeuLysGlyAspValAsn 161
QY 487 AAGGCGCTGAAGCTGAAGGACGGCGCGCCACTACCTGCTGGAGTTCAAGTCCATCTACATG 546
Db 162 MetAlaLeuLeuLeuGluGlyGlyIleTyrArgCysAspPheLysThrThrTyrLys 181
QY 547 GCCAAGAGCCCGTGCAGCTCCCGGCTACTACTAGTGGACACCAAGCTGGACATCAC 606
Db 182 AlalysLysPheValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeu 201
QY 607 TCCACACAGGAGACTACCATCTGTTGAGGAGTACGAGGCGCACCGAGGGCGCGCCAC 666

QY 241 GACTACAAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAG 300
 Db 84 AspTyrPheLysGlnThrPheProGluGlyTyrPheTrpGluArgSerMetThrTyrGlu 103
 QY 301 GAGCGCGGTGGCGACCGTGAACGACGACTCTCCCTGCGAGGCGC-----TGC 351
 Db 104 AspGlnGlyIleCysIleAlaThrAsnAspIleThrMetMetGluGlyValAspAspCys 123
 QY 352 TTCATCTACAAGTCAAGTTCATCGCGTGAATTCCTCCGAGCGCGCGTGAATGCGAG 411
 Db 124 PheAlaTyrLysIleArgPheAspGlyValAsnPheProAlaAsnGlyProValMetGln 143
 QY 412 AAGAAGACCATCGGTGGAGCGCTCCACGAGCGCGCTGTACCCCGCGAGCGGTGCTG 471
 Db 144 A-glyThrLeuLysTrpGluProSerThrGluIleMetTyrAlaArgAspGlyValLeu 163
 QY 472 AGGGCGAGATCCACAGCGCGCTGAAGCTGAGGAGCGCGCGCTACTCTGTTGAGTTC 531
 Db 164 LysGlyAspValAsnMetAlaLeuLeuLeuGluGlyGlyIleTyrArgCysAspPhe 183
 QY 532 AAGTTCATCTACATGCCAAGAGCGCGTGCAGCTGCGCGCTACTACTAGTGTGACACC 591
 Db 184 LysThrThrTyrLysAlaLysLysValValArgLeuProAspTyrHisPheValAspHis 203
 QY 592 AAGTGGACATCACTCCACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCAC 651
 Db 204 A-glyLeuIleValSerHisAspLysAspTyrAsnLysValLysLeuHisGluHisala 223
 QY 652 GAGGCGCGCACCATCCTG 669
 Db 224 GluAlaArgHisGlyLeu 229
 RESULT 14
 Q8T6U0 PRELIMINARY; PRT; 236 AA.
 AC Q8T6U0
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Green fluorescent protein.
 OS Dendronephthya sp. SSAL-2002.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
 OC Nephthedeae; Dendronephthya.
 OX NCBI_TaxID=191210;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21927629; PubMed=11929996;
 RA Labas Y.A., Gurskaya N.G., Yanushovich Y.G., Fradkov A.F.,
 RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
 RT "Diversity and evolution of the green fluorescent protein family."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
 DR EMBL; AF420591; AAM10625.1; -
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP; 1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 236 AA; 26840 MW; CBI707CF9334A90 CRC64;
 Alignment Scores:
 Pred. No.: 1.34e-40 Length: 236
 Score: 657.00 Matches: 119
 Percent Similarity: 72.69% Conservative: 38
 Best Local Similarity: 55.09% Mismatches: 59
 Query Match: 53.11% Indels: 0
 DB: 5 Gaps: 0
 US-10-081-864-14 (1-678) x Q8T6U0 (1-236)
 QY 16 AACGTATCAACCGATTCATCGCTTCAAGTGGGAGCGCGTGAACCGGCAC 75
 Db 2 AsnLeuIleLysGluAspMetArgValLysValHisMetGluGlyAsnValAsnGlyHis 21

QY 76 GACTTCAGATCGAGCGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 135
 Db 22 AlaPheValIleGluGlyGlyArgProTyrGluGlyThrGlnThrLeuAsn 41
 QY 136 TTCAAGGTGACCAAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
 Db 42 LeuThrValLysGluGlyAlaProLeuProPheSerTyrAspIleLeuThrAlaLeu 61
 QY 196 CAGTACGGCTCCAAGGTGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 255
 Db 62 HisTyrGlyAsnArgValPheThrGluTyrProAlaAspIleThrAspTyrPheLysGln 81
 QY 256 TCCTTCCCGAGGCTCAAGTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 315
 Db 82 SerPheProGluGlyTyrSerTrpGluArgThrMetThrTyrGluAspLysGlyIleCys 101
 QY 316 ACCTGACCGGAGGCTCTCCCTGCGAGGCGGCTTCTATCTACAGGTGAAGTTCATC 375
 Db 102 ThrIleArgSerAspIleSerLeuGluGlyAspCysPhePheGlnAsnIleArgPheAsn 121
 QY 376 GGCGTGAACCTCCCTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
 Db 122 GlyMetAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
 QY 436 TCCACCGAGCGCTGTACCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
 Db 142 SerThrGluLysLeuHisValArgAspGlyLeuLeuValGlyAsnIleAsnMetAlaLeu 161
 QY 496 AAGCTGAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
 Db 162 LeuLeuGluGlyGlyHisTyrLeuCysAspPheLysThrThrTyrLysAlaLysLys 181
 QY 556 CCGCTGCGAGTCCCGCGGCTACTACTAGTGACACCAAGCGGCGGCGGCGGCGGCGG 615
 Db 182 ValValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeuSerAsnAsp 201
 QY 616 GAGGACTACACCATCGTGGAGCAGTACGAGCGGCGGCGGCGGCGGCGGCGGCGG 663
 Db 202 SerAspTyrAsnLysValLysLeuTyrGluHisGlyValAlaArgTyr 217
 RESULT 15
 Q9U6Y3
 ID Q9U6Y3 PRELIMINARY; PRT; 266 AA.
 AC Q9U6Y3
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Fluorescent protein FP484.
 OS Clavularia sp.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
 OC Clavulariidae; Clavularia.
 OX NCBI_TaxID=86521;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99436614; PubMed=10504696;
 RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
 RA Markelov M.L., Lukyanov S.A.;
 RT "Fluorescent proteins from nonbioluminescent Anthozoa species."
 RL Nat. Biotechnol. 17:969-973(1999).
 DR EMBL; AF168424; AAF03374.1; -
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP-like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP; 1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 266 AA; 30450 MW; B4E97406E2708854 CRC64;
 Alignment Scores:
 Pred. No.: 1.93e-40 Length: 266
 Score: 655.00 Matches: 119
 Percent Similarity: 73.27% Conservative: 40

Best Local Similarity:	54.84%	Mismatches:	58
Query Match:	52.95%	Indels:	0
DB:	5	Gaps:	0
 US-10-081-864-14 (1-678) x Q9U6Y3 (1-266)			
QY	19	GTCAATCACCAGGTTTCATCGCTTCAAGTGCGCATGGAGGCACCCGTGAACGCCACGAG	78
Dd	45	VallLeyPrAsPmetLyslelLysLeuLysMetGlUGlyAsnValAenGliyHiSaIia	64
QY	79	TTCGAGATCAGGGGAGGGCGGCGCCCTTACGAGGGCCACACACCGTGAAGTTG	138
Dd	65	PheValileGuLiUGlyUGlyUGlyProLyArSpGLyThrHisThrLeuaenLeu	84
QY	139	AAGGTGACCAAGGGGGCCCCCTCCCTTCCTCGGACATCTGTGCCCCCAGTTCCAG	198
Dd	85	GUuValLysGLUGlyAlaProLeuProPheSerTyRAsPleLeuSerAsnaIaPheGln	104
QY	199	TACGGTCCAAGGTGTAGTGAAGCACCCCCCGACATCCCACATACAAGAAGCTGCC	258
Dd	105	TyrGLyAsnArgAlalaLeuThrLysTy:ProAsPaspIleaIaaSPtyrPhelysGlnSer	124
QY	259	TTCGCCGAGGCTTCAAGTGGAGGGGTGATGAATTTCGAGACGGCGCGGTGGCGACC	318
Dd	125	PheProGLUGlyTy:SerTrpGLuArgThrMetThrPheGLuAsPLysGLyleValllys	144
QY	319	GTGACCCAGGACTCTCCCTCGAGACGGCTGCTTCATCTACAAGGTGAAGTTTCATCGC	378
Dd	145	ValLysSerAsPIleSeRMeTGLuAsPSer:PheIleTyrgluileargPheasPgily	164
QY	379	GTGAAGTCCCTCCGACGGCCCGTGTATGACAGAAGAAGACCATGGCTGGAGAGCCTCC	438
Dd	165	MetAnPheProAsnGLyProValMetGLnLysLysThrLeuLysTrpGLuProSer	184
QY	439	ACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGGGAGATCCACAAGGCCCTGAAG	498
Dd	185	ThrGLulleMetTy:ValArgasPgilyValLeuVa:GLyAsPIleSerHisserLeuLeu	204
QY	499	CTGAAGACGGCGGCCACTACTGGTGGAGGTTCAGTTCATCTACATGCCCCAAGAGCCC	558
Dd	205	LeuGLUGlyGLyGLyHIStyrArgCyAsPPheLysSerIleTyryeAlalysLysVal	224
QY	559	GTGACGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACTCCOCACACGAG	618
Dd	225	ValLysLeuProAsPTyHisPheValAsPHiARgilleGuilleuasnHisasplys	244
QY	619	GACTACACCATCTGGAGAGTAGTACGAGCGCACCGAGGGCGGCCAACCACTG	669
Dd	245	AspyrAsnLysValThrLeuTyrgLuAsnAlaValaIaArgTy:SerLeu	261

Search completed: July 29, 2004, 14:36:14
Job time : 44.3863 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 08:30:38 ; Search time 470.142 Seconds
(without alignments)
6388.441 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 707
Sequence: 1 ggtacgcctccctgctgac.....ctggccacactgaagctt 707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	707	100.0	707	6 AAD46286	Aad46286 Anemonia
2	694.4	98.2	1398	7 ACA61028	ACA61028 DNA encod
3	691.2	97.8	699	6 AAD46287	Aad46287 Anemonia
4	689.6	97.5	1404	7 ACA61027	ACA61027 DNA encod
5	676.8	95.7	696	3 AAAS0885	AAAS0885 A. sulcat
6	440.2	62.3	696	4 AAD03616	Aad03616 Anemonia
7	440.2	62.3	696	6 AAD46279	Aad46279 Anemonia
8	440	62.2	696	7 AAAS0884	AAAS0884 Anemonia
9	436.8	61.8	757	7 ABA00799	Abao0799 Wild type
10	436.8	61.8	955	3 AAAS0883	AAAS0883 Anemonia
11	435.2	61.6	767	7 ABA00800	Abao0800 Kindling
12	433.6	61.3	767	7 ABA00801	Abao0801 Kindling
13	417.8	59.1	687	6 ABL41173	ABL41173 H. crisp
14	417.8	59.1	687	6 ABL41172	ABL41172 H. crisp
15	417.8	59.1	955	6 AAI69492	AAI69492 A. sulcat
16	417.8	59.1	1396	6 ABL41174	ABL41174 Cr-449-ta
17	417.8	59.1	1396	7 ACA61024	ACA61024 DNA encod
18	417.8	59.1	1424	6 ABL41175	ABL41175 Cr-449-ta
19	417.8	59.1	1424	7 ACA61025	ACA61025 DNA encod
20	315	44.6	675	6 AAD46281	Aad46281 Discosoma
21	313.4	44.3	678	6 AAD46282	Aad46282 Discosoma
22	313.4	44.3	678	6 AAD28209	Aad28209 Discosoma
23	313.4	44.3	681	9 ADC24130	Adc24130 Discosoma

24	312.6	44.2	678	9 ADC24132	Adc24132 Discosoma
25	310.2	43.9	678	6 AAD46278	Aad46278 Discosoma
26	310.2	43.9	678	6 AAD28207	Aad28207 Discosoma
27	310.2	43.9	678	9 ADC24128	Adc24128 Discosoma
28	310.2	43.9	695	3 AAA48743	AAA48743 Humanised
29	308.6	43.6	678	6 AAD28208	Aad28208 Discosoma
30	308.6	43.6	681	9 ADC24134	Adc24134 Discosoma
31	308.6	43.6	4692	6 AAL47954	Aal47954 Modified
32	308.6	43.6	4692	7 ACC44640	Acc44640 Vector pD
33	308.6	43.6	6893	9 ADE24111	Ade24111 Proviral
34	308.6	43.6	7910	4 AAD09979	Aad09979 pBit(dHSP
35	308.6	43.6	9320	6 ABS56664	ABs56664 Plasmid p
36	307	43.4	723	7 ABZ22476	Abz22476 Mammalian
37	307	43.4	5436	4 RAD10003	Rad10003 Plasmid p
38	299.2	42.3	6985	7 ACA55359	AcA55359 Transform
39	286.2	40.5	681	4 AAH47656	Aah47656 Anthozoan
40	284	40.2	835	6 ABA00244	ABa00244 C. Gigant
41	282.4	39.9	681	6 ABA00245	ABa00245 C. Gigant
42	271.8	38.4	678	3 AAA48750	AAA48750 Humanised
43	271.8	38.4	678	3 AAA48748	AAA48748 Humanised
44	271.8	38.4	678	6 RAD46280	Rad46280 Discosoma
45	270.8	38.3	760	7 ABA00806	ABa00806 Multiple

ALIGNMENTS

RESULT 1					
AAD46286					
ID	AAD46286 standard; DNA; 707 BP.				
XX					
AC	AAD46286;				
XX					
DT	27-DEC-2002 (first entry)				
XX					
DE	Anemonia sulcata asFP595 non-aggregating mutant DNA, FP7-NA.				
XX					
KW	Fluorescent protein; chromoprotein; protease cleavage assay; filter;				
KW	fluorescence activated cell sorting application; fluorescent timer;				
KW	biosensor; fluorescence resonance energy transfer application; FRET;				
KW	colouring agent; recombinant DNA application; analyte detection assay;				
KW	sunscreen; second messenger detector; asFP595 protein; NFP-7; Gene;				
mutant; ds.					
XX					
OS	Anemonia sulcata.				
XX	Synthetic.				
XX					
FT	Key				
CDS	Location/Qualifiers				
FT	7..702				
FT	/*tag= a				
FT	/product= "asFP595 non-aggregating mutant protein #1"				
FT	/note= "No start codon"				
FT	/partial				
XX					
WO	WO200268459-A2.				
XX					
06-SEP-2002.					
XX					
20-FEB-2002; 2002WO-US005749.					
XX					
21-FEB-2001; 2001US-0270983P.					
04-DEC-2001; 2001US-00006922.					
XX					
(CLON-) CLONTECH LAB INC.					
XX					
Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;					
XX					
WPI; 2002-691654/74.					
P-PSDB; AA28841.					
XX					
New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant					
of an aggregating Chidarian chromo- or fluorescent protein or mutant for					
analyte detection assays or fluorescence activated cell sorting					

PT applications.
PS Claim 4; Page 77; 80pp; English.
XX
XX The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays, as colouring agents, as markers in
CC recombinant DNA applications, as sunscreens or filters, in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC sequence is Anemonia sulcata asFP595 non-aggregating mutant DNA of the
XX invention
XX
SQ Sequence 707 BP; 137 A; 264 C; 202 G; 104 T; 0 U; 0 Other;
Query Match 100.0%; Score 707; DB 6; Length 707;
Best Local Similarity 100.0%; Pred. NO. 1.2e-108;
Matches 707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATCCGCTCCCTGCTGACCGGACCATGCTTTCAGGACCACTCCAGGCGCACCGTG 60
Db 1 GGATCCGCTCCCTGCTGACCGGACCATGCTTTCAGGACCACTCCAGGCGCACCGTG 60
QY 61 AACGGCCACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTTCGAGGGCACCCAG 120
Db 61 AACGGCCACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTTCGAGGGCACCCAG 120
QY 121 GAGATGAAGATCGAGGTGATCGAGGGCGGCCCTTCCCTTCCATCTCCTGTCC 180
Db 121 GAGATGAAGATCGAGGTGATCGAGGGCGGCCCTTCCCTTCCATCTCCTGTCC 180
QY 181 ACTCTCTGCATGTACGGCTTCAAGGCTTTCATCAAGTACGTTCGGGCATCCCGCTAC 240
Db 181 ACTCTCTGCATGTACGGCTTCAAGGCTTTCATCAAGTACGTTCGGGCATCCCGCTAC 240
QY 241 TTCAAGCAGTCCCTCCCGAGGGCTTCACTGGAGGGCGACCACTACGAGGACGGC 300
Db 241 TTCAAGCAGTCCCTCCCGAGGGCTTCACTGGAGGGCGACCACTACGAGGACGGC 300
QY 301 GGCTTCTGTGACCGGCCACCAAGGACACCTCCCTGGAAGGGCGACCTCTGTGTACAAGGTG 360
Db 301 GGCTTCTGTGACCGGCCACCAAGGACACCTCCCTGGAAGGGCGACCTCTGTGTACAAGGTG 360
QY 361 AAGATCTGGGCAACACTTCCCGCGGACCGCCCGTGTATGAGAACAGGCGCGCGC 420
Db 361 AAGATCTGGGCAACACTTCCCGCGGACCGCCCGTGTATGAGAACAGGCGCGCGC 420
QY 421 TGGAGAGCCCTCCACCGAGATGCTGTACAGGTGAGCGGCTGTGCGCGCCAGTCCCTG 480
Db 421 TGGAGAGCCCTCCACCGAGATGCTGTACAGGTGAGCGGCTGTGCGCGCCAGTCCCTG 480
QY 481 ATGCGCTTGAAGTCCCGCGGCTGCGCACTGACCTGCGCACTGCGCACTGCGCACTACCGC 540
Db 481 ATGCGCTTGAAGTCCCGCGGCTGCGCACTGACCTGCGCACTGCGCACTACCGC 540
QY 541 TCCAAAGAGCCCGCTCCCGCTTGAAGATCCCGGCTTCCATCTCGAGGACCAACCGCATC 600
Db 541 TCCAAAGAGCCCGCTCCCGCTTGAAGATCCCGGCTTCCATCTCGAGGACCAACCGCATC 600
QY 601 GAGATCTTGAAGAGGTGGAGAGGGCAAGTGTGTACAAGCAGTACGAGGCGCGCGTGGGC 660
Db 601 GAGATCTTGAAGAGGTGGAGAGGGCAAGTGTGTACAAGCAGTACGAGGCGCGCGTGGGC 660
QY 661 CGCTACTTGGACCGCGCCCTTCCAGCTGGGCGACACTGAGCTT 707
Db 661 CGCTACTTGGACCGCGCCCTTCCAGCTGGGCGACACTGAGCTT 707
RESULT 2
ACA61028
ID ACA61028 standard; DNA; 1398 BP.

XX
AC ACA61028;
XX
DT 09-JUL-2003 (first entry)
XX
DE DNA encoding chromoprotein AsRed-35-5D-tandem.
XX
KW Chromo/fluorescent domain; labeled fusion protein;
KW site-specific gene modification; chromoprotein; colouring agent;
KW food composition; pharmaceutical; cosmetic; AsRed-35-5D-tandem; gene; ds.
XX
OS Anthozoa.
XX
FH Key Location/Qualifiers
FT CDS 1..1398
FT /tag= a
FT /product= "AsRed-35-5D-tandem"
XX
XX WO2003031590-A2.
XX
PD 17-APR-2003.
XX
PF 10-OCT-2002; 2002WO-US032560.
XX
PR 12-OCT-2001; 2001US-00976673.
PR 11-FEB-2002; 2002US-0356225P.
PR 22-MAY-2002; 2002US-0383336P.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukanov SA;
XX
XX WPI; 2003-381709/36.
XX P-PSDB; ABU09925.
XX
PT New nucleic acid encoding polypeptide products having at least two linked
PT chromo/fluorescent domains, useful for generating transgenic plants or
PT animals or site-specific gene modifications in cell lines.
XX
PS Disclosure; Fig 5; 68pp; English.
XX
XX The invention describes a nucleic acid encoding a polypeptide product
CC comprising a first and a second chromo/fluorescent domain, optionally
CC joined by a linking domain. The first and second chromo/fluorescent
CC domains associate with each other under intracellular conditions so that
CC the encoded polypeptide assumes a tertiary structure. The nucleic acid
CC and the protein are useful in producing labeled fusion proteins that have
CC a precise and predictable signal to fusion partner ratio. The nucleic
CC acid may also be used in generating transgenic, non-human plants or
CC animals or site-specific gene modifications in cell lines. The
CC chromoproteins may be used as colouring agents, as a food composition, in
CC pharmaceuticals or cosmetics, as labels in analyte detection assays or as
CC selectable markers in recombinant DNA applications. This sequence encodes
CC chromo/fluorescent domain fusion protein AsRed-35-5D-tandem
XX
SQ Sequence 1398 BP; 270 A; 526 C; 400 G; 202 T; 0 U; 0 Other;
Query Match 98.2%; Score 694.4; DB 7; Length 1398;
Best Local Similarity 99.9%; Pred. No. 1.5e-106;
Matches 695; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 7 GCCTCCCTCTGTCGACGAGACCATGCCCTTTCAGGACCACTTCGAGGCGACCGTGAACGGC 66
Db 1 GCCTCCCTCTGTCGACGAGACCATGCCCTTTCAGGACCACTTCGAGGCGACCGTGAACGGC 60
QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCGAGGCGACCCAGGAGATG 126
Db 61 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCGAGGCGACCCAGGAGATG 120
QY 127 AAGATCGAGGTGATCGAGGGCGGCCCTTCCCTTCCATCTTCCATCTCTGTCCACCTCC 186
Db 121 AAGATCGAGGTGATCGAGGGCGGCCCTTCCCTTCCATCTTCCATCTCTGTCCACCTCC 180

QY 187 TGCATGTACGGTCCAAAGGCTTCATCAAGTACGTGTCGGGATCCCGACTACTTCAAG 246
DB 181 TGCATGTACGGTCCAAAGGCTTCATCAAGTACGTGTCGGGATCCCGACTACTTCAAG 240
QY 247 CAGTCCCTCCCGAGGGCTTCACTGGAGGCGACCACTAGAGGAGCGGGCTTC 305
DB 241 CAGTCCCTCCCGAGGGCTTCACTGGAGGCGACCACTAGAGGAGCGGGCTTC 300
QY 307 CTGACCGGCCACAGGACACCTCCCTGACGCGGACTGCTGGTGTAAGGTGAAGATC 366
DB 301 CTGACCGGCCACAGGACACCTCCCTGACGCGGACTGCTGGTGTAAGGTGAAGATC 360
QY 367 CTGGGCAACAATTCCTCCCGGACGCGCCCGGTGATGACAGAACAGGCGCGGCTGGAG 426
DB 361 CTGGGCAACAATTCCTCCCGGACGCGCCCGGTGATGACAGAACAGGCGCGGCTGGAG 420
QY 427 CCTCCACCGAGATCGTACAGAGTGGACGCGGCTGCGGCGGACGTCCTGATGCGC 486
DB 421 CCTCCACCGAGATCGTACAGAGTGGACGCGGCTGCGGCGGACGTCCTGATGCGC 480
QY 487 CTGGAGTGCCTCCCGGCGTGCACCTGACCTGACCTGACACCACTACCGTCCCAAG 546
DB 481 CTGGAGTGCCTCCCGGCGTGCACCTGACCTGACCTGACACCACTACCGTCCCAAG 540
QY 547 AAGCCCGCTCCGCGCTGAAGATCCCGGCTTCCACTTCGAGGACCAAGGATCGAGATC 606
DB 541 AAGCCCGCTCCGCGCTGAAGATCCCGGCTTCCACTTCGAGGACCAAGGATCGAGATC 600
QY 607 CTGGAGGAGTGGAGAGGGCAAGTGTACAGAGTACGAGGCGCGGCTGGCGGCTTAC 666
DB 601 CTGGAGGAGTGGAGAGGGCAAGTGTACAGAGTACGAGGCGCGGCTGGCGGCTTAC 660
QY 667 TCGGACGCGCGCGCTCCAAAGTGGGCGCAACTGA 702
DB 661 TCGGACGCGCGCGCTCCAAAGTGGGCGCAACTGA 696

RESULT 3

AAD46287
ID AAD46287 standard; DNA; 699 BP.
XX
AC AAD46287;
XX
DT 27-DEC-2002 (first entry)
XX
DE Anemonia sulcata asFP595 non-aggregating mutant DNA, Asred M35-5D.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunsreen; second messenger detector; asFP595 protein; NFP-7; gene;
KW mutant; ds.
XX
OS Anemonia sulcata.
OS Synthetic.
XX
FH Key Location/Qualifiers
CDS 1..699
FT /*tag= a
FT /product= "asFP595 non-aggregating mutant protein"
XX
XX WO200268459-A2.
XX
XX 06-SEP-2002.
XX
XX 20-FEB-2002; 2002WO-US005749.
XX
XX 21-FEB-2001; 2001US-0270983P.
XX
XX 04-DEC-2001; 2001US-00006922.
XX
XX (CLON-) CLONTECH LAB INC.

PI Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX WPI; 2002-691654/74.
DR P-PSDB; AAE28842.
XX
XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX of an aggregating Chidarian chromo- or fluorescent protein or mutant for
XX analyte detection assays or fluorescence activated cell sorting
XX applications.
XX
PS Claim 4; Fig 13; 80pp; English.
XX
CC The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays, as colouring agents, as markers in
CC recombinant DNA applications, as sunsreens or filters, in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC sequence is Anemonia sulcata asFP595 non-aggregating mutant DNA of the
CC invention. Note: This sequence is stated to be the same as that shown as
CC SEQ ID NO:23 in the sequence listing. However this sequence contains
CC additional bases at its 5' end
XX
SQ Sequence 699 BP; 137 A; 261 C; 200 G; 101 T; 0 U; 0 Other;
Query Match 97.8%; Score 691.2; DB 6; Length 699;
Best Local Similarity 99.6%; Pred. No. 5.2e-106;
Matches 693; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
DB 4 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 63
QY 67 CACTACTTCAAGTGCACCGCAAGGGCGAGGGCAACCCCTTCAGAGGCGACCCAGGAGATG 126
DB 64 CACTACTTCAAGTGCACCGCAAGGGCGAGGGCAACCCCTTCAGAGGCGACCCAGGAGATG 123
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTCTGCACCTCC 186
DB 124 AAGATCGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTCTGCACCTCC 183
QY 187 TGCATGTACGGCTCCAAAGGCGCTTCATCAAGTACGTGTCGGGATCCCGGACTACTTCAAG 246
DB 184 TGCATGTACGGCTCCAAAGGCGCTTCATCAAGTACGTGTCGGGATCCCGGACTACTTCAAG 243
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGGCGGACCAACCACTACGAGGCGGGCTTC 306
DB 244 CAGTCCCTCCCGAGGGCTTCACCTGGAGGCGGACCAACCACTACGAGGCGGGCTTC 303
QY 307 CTGACCGGCCACAGGACACCTCCCTGAGCGGCGACTGCTGGTGTAAGGTGAAGATC 366
DB 304 CTGACCGGCCACAGGACACCTCCCTGAGCGGCGACTGCTGGTGTAAGGTGAAGATC 363
QY 367 CTGGGCAACAATTCCTCCCGGCGGCGCTGATGAGAGAACAGGCGCGGCGCTGGAG 426
DB 364 CTGGGCAACAATTCCTCCCGGCGGCGCTGATGAGAGAACAGGCGCGGCGCTGGAG 423
QY 427 CCTCCACCGAGATCGTGTACAGGTGAGCGGCTGCTGCGGCGGCGAGTCCCTGATGGCC 486
DB 424 CCTCCACCGAGATCGTGTACAGGTGAGCGGCTGCTGCGGCGGCGAGTCCAGATGGCC 483
QY 487 CTGGAGTGCCTCCCGGCGGCTGCGGCGCTGACCTGCGGCGGCGGCGGCGCTCCCAAG 546
DB 484 CTGGAGTGCCTCCCGGCGGCTGCGGCGCTGACCTGCGGCGGCGGCGGCGCTCCCAAG 543
QY 547 AAGCCCGCTCCGCGCTGAAGATCCCGGCTTCACCTTCGAGGACCAAGGATCGAGATC 606
DB 544 AAGCCCGCTCCGCGCTGAAGATCCCGGCTTCACCTTCGAGGACCAAGGATCGAGATC 603
QY 607 CTGGAGGAGTGGAGAGGGCAAGTGTCTACAAGCAGTACGAGGCGGCGGCTGGCGGCTTAC 666

Db 604 CTGAGGAGGTGGAGAGGCAAGTGTCTACAGCAGTACGAGGCGCGCTGGGCGGTAC 663
 Qy 667 TGGCAGCGCGCCCTCCAGCTGGGCGCAACTGA 702
 Db 664 TGGCAGCGCGCCCTCCAGCTGGGCGCAACTGA 699

RESULT 4
 ID ACA61027 standard; DNA; 1404 BP.
 XX
 AC ACA61027;
 DT 09-JUL-2003 (first entry)
 XX
 DE DNA encoding chromoprotein AsRed-35-5NA-tandem.
 XX
 KW Chromo/fluorescent domain; labeled fusion protein;
 KW site-specific gene modification; chromoprotein; colouring agent;
 KW food composition; pharmaceutical; cosmetic; AsRed-35-5NA-tandem; gene;
 KW ds.
 XX
 OS Anthozoa.

XX Key Location/Qualifiers
 PH 1..1404
 FT /*tag= a
 FT /product= "AsRed-35-5NA-tandem"
 FT
 FT WO2003031590-A2.
 XX
 XX 17-APR-2003.
 XX
 XX 10-OCT-2002; 2002WO-US032560.
 XX
 XX 12-OCT-2001; 2001US-00976673.
 PR 11-FEB-2002; 2002US-0356225P.
 PR 22-MAY-2002; 2002US-0383336P.
 XX
 XX (CLON-) CLONTECH LAB INC.
 PA
 XX Lukyanov SA;
 XX
 XX WPI; 2003-381709/36.
 DR P-PSDB; ABU09924.
 XX
 XX New nucleic acid encoding polypeptide products having at least two linked
 PT chromo/fluorescent domains, useful for generating transgenic plants or
 PT animals or site-specific gene modifications in cell lines.
 XX
 XX Disclosure; Fig 4; 68pp; English.
 PS
 XX The invention describes a nucleic acid encoding a polypeptide product
 CC comprising a first and a second chromo/fluorescent domain, optionally
 CC joined by a linking domain. The first and second chromo/fluorescent
 CC domains associate with each other under intracellular conditions so that
 CC the encoded polypeptide assumes a tertiary structure. The nucleic acid
 CC and the protein are useful in producing labeled fusion proteins that have
 CC a precise and predictable signal to fusion partner ratio. The nucleic
 CC acid may also be used in generating transgenic, non-human plants or
 CC animals or site-specific gene modifications in cell lines. The
 CC chromoproteins may be used as colouring agents, as a food composition, in
 CC pharmaceuticals or cosmetics, as labels in analyte detection assays or as
 CC selectable markers in recombinant DNA applications. This sequence encodes
 CC chromo/fluorescent domain fusion protein AsRed-35-5NA-tandem
 XX
 XX Sequence 1404 BP; 274 A; 526 C; 402 G; 202 T; 0 U; 0 Other;
 SQ

Query Match 97.5%; Score 699.6; DB 7; Length 1404;
 Best Local Similarity 99.4%; Pred. NO. 9.2e-106;
 Matches 692; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 7 GCCTCCCTGCTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGCGCGTGAACGCGC 66

Db 4 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGCGTGAACGCGC 63
 Qy 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGCGCGTGAACGAGATG 126
 Db 64 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGCGCGTGAACGAGATG 123
 Qy 127 AAGATCGAGAGTGCATCGAGGCGCGCGCTGCGCTTCGCTTCACATCTGTCTCCACCTCC 186
 Db 124 AAGATCGAGAGTGCATCGAGGCGCGCGCTGCGCTTCGCTTCACATCTGTCTCCACCTCC 183
 Qy 187 TGCATGTACGGCTCAAGGCGCTTCATCAAGTACGTGTCCGCGTACCTCCGAGTACTTCAAG 246
 Db 184 TGCATGTACGGCTCAAGGCGCTTCATCAAGTACGTGTCCGCGTACCTCCGAGTACTTCAAG 243
 Qy 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGCGGCGCACCACTACGAGGCGGCGGTTC 306
 Db 244 CAGTCCCTCCCGAGGCGCTTCACCTGGAGCGGCGCACCACTACGAGGCGGCGGTTC 303
 Qy 307 CTGACCGCGCCACCGAGACACCTCCCTCGAGCGGCGACTGCTGTGTGTACAAGGTGAAGATC 366
 Db 304 CTGACCGCGCCACCGAGACACCTCCCTCGAGCGGCGACTGCTGTGTGTACAAGGTGAAGATC 363
 Qy 367 CTGGGCAACAACTTCCCGCGCGCGCGGTGTATGATGAGAACAGGCGCGCGTGGGAG 426
 Db 364 CTGGGCAACAACTTCCCGCGCGCGCGGTGTATGATGAGAACAGGCGCGCGTGGGAG 423
 Qy 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGCGTGTGCGCGGCGCACTCCCTGATGGCC 486
 Db 424 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGCGTGTGCGCGGCGCACTCCACGATGGCC 483
 Qy 487 CTGAGTGCCTCCCGCGCGGTGCGCACTGACCTGCGCGCGCGCACTCCACGATGGCG 546
 Db 484 CTGAGTGCCTCCCGCGCGGTGCGCACTGACCTGCGCGCGCGCACTCCACGATGGCG 543
 Qy 547 AAGCCCGCTCCCGCGCTGAAGATGCGCGGCTTCACCTTCGAGGACCAACCGATCGAGATC 606
 Db 544 AAGCCCGCTCCCGCGCTGAAGATGCGCGGCTTCACCTTCGAGGACCAACCGATCGAGATC 603
 Qy 607 CTGAGGAGGTGGAGAAAGGGCAAGTGTACAAGCAGTACGAGCGCGCGTGGGCGCGTAC 666
 Db 604 CTGAGGAGGTGGAGAAAGGGCAAGTGTACAAGCAGTACGAGCGCGCGTGGGCGCGTAC 663
 Qy 667 TGGCAGCGCGCGCGCTCCAAAGCTGGGCGCAACTGA 702
 Db 664 TGGCAGCGCGCGCGCTCCAAAGCTGGGCGCAACTGA 699

RESULT 5
 ID AAA50885 standard; cDNA; 696 BP.
 XX
 AC AAA50885;
 DT 08-SEP-2000 (first entry)
 XX
 DE A. sulcata humanised mutant fluorescent protein, Mut1, coding sequence.
 XX
 KW Green fluorescent protein; GFP; Mut1; protein localisation;
 KW fluorescence resonance energy transfer; mutant; ss.
 XX
 OS Anemonia sulcata.
 XX
 XX Key Location/Qualifiers
 PH 1..696
 FT /*tag= a
 FT /partial
 FT /product= "Humanised Mut1"
 FT
 FT WO2000034319-A1.
 XX
 XX 15-JUN-2000.
 PD
 XX

PF 10-DEC-1999; 99WO-US029300.
XX
PR 11-DEC-1998; 98US-00210330.
PR 09-DEC-1999; 99US-00210330.
XX
XX (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Jiang X, Duong T;
PI Zhao X;
XX
XX WPI; 2000-423374/36.
DR P-PSDB; AAY99734.
XX
XX Novel fluorescent proteins from non-bioluminescent Anemonia sulcata,
PT useful for fluorescent labeling and as markers.
XX
XX Claim 8; Page 67; 71pp; English.
XX
XX The present sequence is the coding sequence for humanised mutant
CC fluorescent protein, Mut1, from non-bioluminescent Anemonia sulcata. The
CC fluorescent protein has applications in fluorescent labelling, as
CC fluorescent markers for gene expression and protein localisation studies,
CC and in fluorescence resonance energy transfer (FRET) reactions. The
CC coding sequence may be used as a source or primers and probes for
CC identifying related proteins
XX
SQ Sequence 696 BP; 141 A; 257 C; 195 G; 103 T; 0 U; 0 Other;
Query Match 95.7%; Score 676.8; DB 3; Length 696;
Best Local Similarity 98.3%; Pred. No. 1.3e-103;
Matches 684; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 7 GCTCTCCTGTCAGCGAGACGATGCTTTCAGGACACCATCGAGGGACCGTGAACGCG 66
DB 1 GCCTCTTCTGAAGAAGACCATGCTTTCAGGACACCATCGAGGGACCGTGAACGCG 60
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGACCCAGAGATG 126
DB 61 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGACCCAGAGATG 120
QY 127 AAGATCAGAGTGCATCGAGGGCGGCGCTTTCAGGACACCATCGAGGGACCGTGAACGCG 186
DB 121 AAGATCAGAGTGCATCGAGGGCGGCGCTTTCAGGACACCATCGAGGGACCGTGAACGCG 180
QY 187 TGCATGTAGGCTTCCAGGAGGCTTTCATCAAGTACGCTGCGGACATCCCGACTTCAAG 246
DB 181 TGCATGTAGGCTTCCAGGAGGCTTTCATCAAGTACGCTGCGGACATCCCGACTTCAAG 240
QY 247 CAGTCTCCTCCGAGGGCTTTCAGTGGAGCGCACCCACCTAGAGGACCGCGGCTTC 306
DB 241 CAGTCTTCCCGAGGGCTTTCAGTGGAGCGCACCCACCTAGAGGACCGCGGCTTC 300
QY 307 CTGACCGCCACACGAGACCTTCCCTGGAAGGCGGAGTGGTGTACAGGTGAAGATC 366
DB 301 CTGACCGCCACACGAGACCTTCCCTGGAAGGCGGAGTGGTGTACAGGTGAAGATC 360
QY 367 CTGGGCAACAATTCCCGGCGGAGCGGCGGCTGATGACAGCAAGGCGGCGCTGGGAG 426
DB 361 CTGGGCAACAATTCCCGGCGGAGCGGCGGCTGATGACAGCAAGGCGGCGCTGGGAG 420
QY 427 CCCTCCACCGAGATCGTGTACAGAGTGGAGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 486
DB 421 CCCTCCACCGAGATCGTGTACAGAGTGGAGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 480
QY 487 CTGAGTGCCTCCCGGCGGCTGCGGAGTGGAGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 546
DB 481 CTGAGTGCCTCCCGGCGGCTGCGGAGTGGAGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 540
QY 547 AAGCGCGCTCCCGGCTGAGAGTGGCGGCTTCCACTTGGAGGACCCCGCATCGAGATC 606
DB 541 AAGCGCGCTCCCGGCTGAGAGTGGCGGCTTCCACTTGGAGGACCCCGCATCGAGATC 600
QY 607 CTGGAGGAGTGGAGAGGCGCAAGTGTCTACAGAGTACGAGGCGCGCTGGGCCCTAC 666

DB 601 ATGGAGGAGTGGAGAGGCGCAAGTCTACAGCAGTACGAGCGCCCTGGGCCCTAC 660
QY 667 TGGGAGCGCGCCCTCCAGCTGGGCGCACCACTGA 702
DB 661 TGGGAGCGCGCCCTCCAGCTGGGCGCACCACTAA 696
RESULT 6
AAD03616
ID AAD03616 standard; cDNA; 696 BP.
AC AAD03616;
XX
XX 19-JUN-2001 (first entry)
XX
XX Anemonia sulcata chromo/fluorescent protein, asFP600 (NFP-7) cDNA.
KW Anthozoa; Chromoprotein; fluorescent protein; asFP600; NFP-7; sunsreen;
KW analyte detection assay; selectable marker; recombinant DNA application;
KW biosensor; pH indicator; invivo marker; selective filter; sea anemone;
KW ss.
XX
XX Anemonia sulcata.
XX
XX Key Location/Qualifiers
FH CDS 1..696
FT /*tag= a
FT /product= "Chromo/fluorescent protein, asFP600 (NFP-7)"
FT /note= "CDS does not include stop codon"
FT /partial
XX
PN WO200127150-A2.
XX
PD 19-APR-2001.
XX
PF 13-OCT-2000; 2000WO-US028477.
XX
XX 14-OCT-1999; 99US-00418529.
PR 15-OCT-1999; 99US-00418517.
PR 15-OCT-1999; 99US-00418522.
PR 19-NOV-1999; 99US-0044338.
PR 19-NOV-1999; 99US-0044341.
PR 09-DEC-1999; 99US-00457556.
PR 09-DEC-1999; 99US-00457898.
PR 09-DEC-1999; 99US-00458144.
PR 09-DEC-1999; 99US-00458477.
PR 10-DEC-1999; 99WO-US029405.
PR 14-JUN-2000; 2000US-0211607P.
PR 14-JUN-2000; 2000US-0211609P.
PR 14-JUN-2000; 2000US-0211626P.
PR 14-JUN-2000; 2000US-0211627P.
PR 14-JUN-2000; 2000US-0211687P.
PR 14-JUN-2000; 2000US-0211766P.
PR 14-JUN-2000; 2000US-0211860P.
PR 14-JUN-2000; 2000US-0211888P.
PR 14-JUN-2000; 2000US-0212070P.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Tersikh A;
XX
XX WPI; 2001-266409/27.
DR P-PSDB; AAE00376.
XX
XX An Anthozoa chromo- or fluorescent protein (P1) present in an environment
PT other than its natural environment, useful as a label in analyte
PT detection assays and as a selectable marker in recombinant DNA
PT applications.
XX
XX Claim 13; Fig 7; 69pp; English.
XX
XX The present sequence is an Anemonia sulcata chromo/fluorescent protein,
CC

DT	XX	27-DEC-2002 (first entry)	
DE	XX	Anemonia sulcata asFP600 (NFP-7) wild-type protein encoding DNA.	
KW	XX	Fluorescent protein; chromoprotein; protease cleavage assay; filter;	
KW	XX	fluorescence activated cell sorting application; fluorescent timer;	
KW	XX	biosensor; fluorescence resonance energy transfer application; FRET;	
KW	XX	colouring agent; recombinant DNA application; analyte detection assay;	
XX	XX	sunscreen; second messenger detector; asFP600 protein; NFP-7; gene; ds.	
OS	XX	Anemonia sulcata.	
PH	XX	Key Location/Qualifiers	
FT	FT	1..696	
FT	FT	/*tag= a	
FT	FT	/product= "asFP600 wild-type protein"	
FT	FT	/transl_except= (pos:202..204, aa:Thr)	
FT	FT	/transl_except= (pos:427..429, aa:Ala)	
FT	FT	/note= "No stop codon"	
FT	FT	/partial	
XX	XX		
PN	XX	W0200268459-A2.	
XX	XX		
PD	XX	06-SEP-2002.	
XX	XX		
XX	XX	20-FEB-2002; 2002WO-US005749.	
XX	XX		
PR	XX	21-FEB-2001; 2001US-0270983P.	
PR	XX	04-DEC-2001; 2001US-00006922.	
XX	XX		
PA	XX	(CLON-) CLOWTECH LAB INC.	
XX	XX		
PI	XX	Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;	
XX	XX		
DR	XX	WPI; 2002-691654/74.	
DR	XX	P-PSDB; AA528834.	
XX	XX		
PT	XX	New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant	
PT	XX	of an aggregating Cnidarian chromo- or fluorescent protein or mutant for	
PT	XX	analyte detection assays or fluorescence activated cell sorting	
PT	XX	applications.	
XX	XX		
PS	XX	Disclosure; Page 71; 80pp; English.	
XX	XX		
CC	XX	The invention relates to nucleic acid molecules encoding non-aggregating	
CC	XX	chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are	
CC	XX	useful in analyte detection assays, as colouring agents, as markers in	
CC	XX	recombinant DNA applications, as sunscreens or filters, in fluorescence	
CC	XX	resonance energy transfer (FRET) applications, as biosensors in	
CC	XX	prokaryotic and eukaryotic cells, in screening assays, as second	
CC	XX	messenger detectors, in fluorescence activated cell sorting applications,	
CC	XX	in protease cleavage assays or as fluorescent timers. The present	
CC	XX	sequence is a DNA encoding Anemonia sulcata asFP600 (NFP-7) wild-type	
CC	XX	protein of the invention	
XX	XX		
SQ	XX	Sequence 696 BP; 195 A; 147 C; 176 G; 178 T; 0 U; 0 Other;	
		Query Match 62.3%; Score 440.2; DB 6; Length 696;	
		Best Local Similarity 77.2%; Pred. No. 2.2e-64;	
		Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0	
Qy	7	GCTTCCTGCTGACCGAGACCATCGCTTCAGGACACCATTCAGGGCACCGGACCGG 66	
Db	4	GCTTCCTTTTAAAGAAGACTATGCCCTTTAAGACGACCATTCAGGGACGGTTAATGCG 63	
Qy	67	CACCTACTTTCAGTGCACCGGACGGGACCGGACCCCTCCAGGGCACCCAGGAGATG 126	
Db	64	CACCTACTTTCAGTGTACAGGNAAGAGAGGGGCAACCCATTTGAGGGGTACGAGGAATG 123	
Qy	127	AAGATCGAGGTGATCGAGGGCGGGCCCCCTCGCCTTCGCTTCCACATCTGTCTCACCTCC 186	
Db	124	AAGATAGAGGTCATCGAGGAGGTTCCATTGCCATTTTCCTTCCACATTTGTCAACGAGT 183	

XX ABA00799;
 XX AC
 XX DT
 XX DE
 XX DE Wild type kindling fluorescent protein, AsFP595 coding sequence.
 XX KW Gene; kindling fluorescent protein; kindling stimulus; movement;
 XX KW labeling; fluorescence resonance energy transfer; FRET;
 XX KW bioluminescence resonance energy transfer; BRET; biosensor;
 XX KW automated screening; ss.
 XX OS Anthozoa sulcata.
 XX FH Key Location/Qualifiers
 XX FT CDS 69..763
 XX FT /*tag= a
 XX FT /product= "AsFP595"
 XX FT
 XX PN WO200296924-A1.
 XX XX 05-DEC-2002.
 XX XX 24-MAY-2002; 2002WO-US016379.
 XX XX 25-MAY-2001; 2001US-0293752P.
 XX XX 11-OCT-2001; 2001US-0329176P.
 XX XX (CLON-) CLONTECH LAB INC.
 XX XX Lukyanov SA, Chudakov D, Lukyanov K;
 XX XX NPT; 2003-156788/15.
 XX XX P-PSDB; AAG79762.
 XX XX Novel nucleic acid that is present in other than its natural environment
 XX XX and that encodes kindling fluorescent protein, is useful in labeling
 XX XX protocols, e.g. labeling proteins, organelles, cells and organisms.
 XX XX Example; Fig 1; 96pp; English.
 XX XX The sequences given in ABA00799-805 encode wild type and mutant kindling
 XX XX fluorescent proteins. The proteins go from a first substantially non-
 XX XX fluorescent or non-fluorescent state to a second fluorescent state upon
 XX XX exposure to a kindling stimulus. The kindling proteins are useful for
 XX XX detecting an entity such as a protein, organelle or cell in a composition
 XX XX such as a cell or a multicellular composition (preferably a multicellular
 XX XX organism), by providing the entity as an entity labeled with the kindling
 XX XX protein, kindling the kindling fluorescent protein label with a kindling
 XX XX stimulus to produce a kindled kindling fluorescent protein label, and
 XX XX exciting the kindled kindling fluorescent protein label with light and
 XX XX detecting any fluorescence from it to detect the entity. The method
 XX XX monitors the movement of the entity. The fluorescent proteins and the
 XX XX cDNA encoding them are useful in labeling protocols, e.g., labeling
 XX XX proteins, organelles, cells and organisms, as biological labels or
 XX XX markers, in protein labeling or tagging applications. The fluorescent
 XX XX kindling proteins are useful as detectable labels, as labels in analyte
 XX XX detection assays, in fluorescence resonance energy transfer (FRET)
 XX XX applications, in bioluminescence resonance energy transfer (BRET)
 XX XX applications, as biosensors in prokaryotic and eukaryotic cells, in
 XX XX applications involving the automated screening of arrays of cells
 XX XX expressing fluorescent reporting groups, in high throughput screening
 XX XX assays, as second messenger detectors, and in fluorescent activated cell
 XX XX sorting assays
 XX XX Sequence 767 BP; 225 A; 163 C; 187 G; 192 T; 0 U; 0 Other;
 XX XX
 XX XX Query Match 61.8%; Score 436.8; DB 7; Length 767;
 XX XX Best Local Similarity 76.7%; Pred. NO. 7.9e-64;
 XX XX Matches 534; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
 XX XX 7 GCTCTCCCTGTCAGCAGACCATGCCCTTCAGGACCACTCGAGGACCGTGAACGGC 66

Db 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTAAGGAGACGGTTAATGGC 131
 QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGGCAACCCCTCGAGGGGCAACCCAGAGATG 126
 Db 132 CACTACTTCAAGTGTACAGGAAAAGAGAGAGGGCAACCCATTTGAGGGGTACGAGGAAATG 191
 QY 127 AAGATCGAGGTGATCGAGGGGGCCCTGCTTGGCTTCCACATCTCTGTCACCTCC 186
 Db 132 AAGATAGAGGTATCGAAGGAGGTCCATGTCATTTGCCCTTCCACATTTTGTCAACGAT 251
 QY 187 TGCATGTACGGCTCCAGGGCTTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
 Db 252 TGTATGTACGGTAGTAAGACCTTTCATCAAGTATGTGTGAGGAATTCCTGACTACTTCAAG 311
 QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCACTACGAGGAGCGGGCTTC 306
 Db 312 CAGTCTTTCCCTGAAGGTTTACTTTGGGAAAGAACCACTACGAGGATGGAGGCTTT 371
 QY 307 CTGACCGCCACACAGGACACCTCCCTGGACGGCGACTGCTGTGTACAAGGTGAAGATC 366
 Db 372 CTTACAGCTCATCAGGACACAAAGCTAGATGAGATTGGCTCGTTTACAAGGTCAAGAT 431
 QY 367 CTGGGCAACACTTCCCGCGAGCGGCCCTGTGTGAGAACAAAGGCGCGCTGGGAG 426
 Db 432 CTTGGTAAATAATTTTCCCTGTGATGGGCCCGTGTGATGAGAACAAAGGAGGATGGGAG 491
 QY 427 CCCTCCACCGAGATCGTGTACGAGGTGGACGGCGTGTGCGCGGCCAGTCCCTCATGGCC 486
 Db 492 CCAGCCACCGAGATAGTTTATGAAGTTGACGGTGTCTGCTGAGACAGTCTTTGATGGCC 551
 QY 487 CTGAGATGCCCGCGGGTGGCCACTGACCTGCCACTGACACCACTACCGTCCCAAG 546
 Db 552 CTTAAGTGCCTCGTGGTGTGCTATCTGACTTGCCATCTCCATCTACTTACAGGTCCAAA 611
 QY 547 AAGCCCGCTCCCGCTGAAAGATGCCCGCTTCCACTTCGAGGACCACTCGAGATC 606
 Db 612 AAACAGTAGTGTGCTTGAAGATGCCAGGATTTCAATTTGAAGATATCCGATCGAGATA 671
 QY 607 CTGAGAGGTGGAGAGGGCAAGTGTCTACAAGCAGTACGAGGCGCGCTGGGCGCTAC 666
 Db 672 ATGAGGAAAGTTGAGAAAGGCAAGTGTCTATAAAGTACGAGCAGTGGGCGAGGTAC 731
 QY 667 TGCACCGCGCCCTCCAGCTGGGCGCACTGA 702
 Db 732 TGTGATGCTGCTCCATCCAAAGCTTGGACATAACTAA 767
 RESULT 10
 AAA50883
 ID AAA50883 standard; cDNA; 955 BP.
 XX AC AAA50883;
 XX DT 08-SEP-2000 (first entry)
 XX DE Anemonia sulcata fluorescent protein, asFP600, coding sequence.
 XX KW Green fluorescent protein; GFP; asFP600; protein localisation;
 XX KW fluorescence resonance energy transfer; ss.
 XX OS Anemonia sulcata.
 XX FH Key Location/Qualifiers
 XX FT CDS 69..767
 XX FT /*tag= a
 XX FT /product= "asFP600 fluorescent protein"
 XX FT
 XX PN WO200034319-A1.
 XX XX 15-JUN-2000.
 XX XX 10-DEC-1999; 99WO-US029300.
 XX XX

PR	11-DEC-1998;	98US-00210330.
PR	09-DEC-1999;	99US-00210330.
XX	(CLON-) CLONTECH LAB INC.	
XX	Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Jiang X, Duong T;	
PI	Zhao X;	
XX	WPI; 2000-423374/36.	
DR	P-PSDB; AAY99733.	
XX	Novel fluorescent proteins from non-bioluminescent Anemonia sulcata,	
PT	useful for fluorescent labeling and as markers.	
XX	Claim 8; Page 63-64; 71pp; English.	
XX	The present sequence is the coding sequence for fluorescent protein,	
CC	asFP600, from non-bioluminescent Anemonia sulcata. The fluorescent	
CC	protein (AAY99733) has applications in fluorescent labelling, as	
CC	fluorescent markers for gene expression and protein localisation studies,	
CC	and in fluorescence resonance energy transfer (FRET) reactions. The	
CC	coding sequence may be used as a source or primers and probes for	
CC	identifying related proteins	
XX		
SQ	Sequence 955 BP; 273 A; 196 C; 216 G; 270 T; 0 U; 0 Other;	
	Query Watch	61.8%; Score 436.8; DB 3; Length 955;
	Best Local Similarity	76.7%; Pred. No. 7.8e-64; Indels 0; Gaps 0;
	Matches 534; Conservative 0; Mismatches 162;	
QY	7 GCCTCCCTGCTGCACGAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGCGC 66	
Db	72 GCTTCTCTTTTAAGAAGACTATGCCCTTTAAAGACGACCATTAAGGGACGGTGAATGGC 131	
QY	67 CACTACTTCAAGTGACCGGCAAGGGGAGGGCAACCCCCTCAGGGCCACCCAGGATG 126	
Db	132 CACTACTTCAAGTGATACGAGAAAAGGAGGGCAACCCATTGAGGGTAGCGAGGAATG 191	
QY	127 AAGATCGAGGTGATCGAGGGGGGGCCCCCTGCGCTTCGCCCTTCACATCTCGTGCACCTCC 186	
Db	192 AAGATAGAGTTCATCGAAGAGAGTCCATTGCCATTTGCCCTTCACATTTGTCAACGAGT 251	
QY	187 TCATGTGTACGGCTCCAAGGCCCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246	
Db	252 TGATGTACGGTAGTGAAGACCTTCATCAAGPATGTGTGAGGAATTCCTGACTACTTCAAG 311	
QY	247 CAGTCCCTCCCGAGGGCTTCACCTGGGAGGGCACCAACCTACGAGGACGGCGGCTTC 306	
Db	312 CAGTCTTCCCTGAGAGTTTTACTTTGSAANAAGAACACAACCTACGAGATGAGGCTTT 371	
QY	307 CTGACGGCCACACGGAACCTTCCTGGACGGCGACTGCCTGGTGTACAAAGTGAAGATC 366	
Db	372 CTTACAGCTCATCAGGACACAAGCCTAGATGGAGATTGCCTCGCTTACAAAGGTCAAGATT 431	
QY	367 CTGGGCAACAATCTCCCGCGAGGGCCCGGTGATGAGAACAAAGCGCGCGCTGGGAG 426	
Db	432 CTTGGTAAATAATTTTCCTGCTGATGGCCCCGTGATGAGAGCAAAAGCAGAAAGTGGAG 491	
QY	427 CCTCCACCGAGATGCTGTACAGGTGACGGCGCTGTGGCGGCCAGTCCCTGATGGCC 486	
Db	492 CCAGCCACCGAGATAGTTTATGAAGTTGACGGTGTCTGCTGGACAGCTCTTTGATGGCC 551	
QY	487 CTGGAGTGCCTGGGGGTGCGCACTGACCTGCCACCTGCACACCACCTACCGTCCAG 546	
Db	552 CTTAAGTGCCTCGTGGTGTGCATCTGATCTGGCCATCTCCATCTACTTTACAGGTCCTCAA 611	
QY	547 AAGCCCGCTCCGCGCTGAAGATGCCGGCTTCACCTTCGAGGACACACCGCATCGAGATC 606	
Db	612 AAACAGCTAGTGGCTTGAGNATGCCAGATTTCAATTTGAAGATCATCGCATCGAGATA 671	
QY	607 CTGAGGAGGTGGAGAGGCAAGTGTCTAAGCAGTACGAGGCGCGCGTGGGCGCGCTAC 666	
Db	672 ATGAGGGAAGTGTGAGAAAGGCAAGTGTCTATAACAGTACGAAGCAGCAGTGGCGCAGTAC 731	

Qy 667 TGGACGCGCCCTTCAAGCTGGGCACAACTGA 702
|||
Db 732 TTGTATGCTGTCCATCAAGCTTGGACATACTAA 767
|||

RESULT 11	
ABAC0800	
ID	ABAC0800 standard; cDNA; 767 BP.
XX	
XX	ABAC0800;
XX	
XX	01-APR-2003 (first entry)
XX	
DE	Kindling fluorescent protein, ASPF595 A148G coding sequence.
XX	
KW	Gene; kindling fluorescent protein; kindling stimulus; movement;
KW	labeling; fluorescence resonance energy transfer; FRET;
KW	bioluminescence resonance energy transfer; BRET; biosensor;
KW	automated screening; ss.
XX	
OS	Anthozoa sulcata.
XX	
XX	
Key	Location/Qualifiers
FT	69..763
FT	/*tag= a
FT	/product= "ASPF595 A148G"
FT	replace(496,C)
FT	/*tag= b
FT	
XX	
XX	WC200296924-A1.
XX	
XX	
PD	05-DEC-2002.
XX	
PF	24-MAY-2002; 2002WO-US016379.
XX	
XX	25-MAY-2001; 2001US-0293752P.
PR	
PR	11-OCT-2001; 2001US-0329176P.
XX	
XX	(CLON-) CLONTECH LAB INC.
PA	
XX	
XX	Lukyanov SA, Chudakov D, Lukyanov K;
PI	
XX	
DR	WPI; 2003-156788/15.
XX	
DR	P-PSDB; AAG79763.
XX	
XX	
PT	Novel nucleic acid that is present in other than its natural env
FT	and that encodes kindling fluorescent protein, is useful in lab
PT	protocols, e.g. labeling proteins, organelles, cells and organ
XX	
PS	Example; Fig 2; 96pp; English.

100

```
CC assays, as second messenger detectors, and in fluorescent activated cell
CC sorting assays
XX
SQ Sequence 767 BP; 225 A; 162 C; 188 G; 192 T; 0 U; 0 Other;

Query Match 61.6%; Score 435.2; DB 7; Length 767;
Best Local Similarity 76.6%; Pred. No. 1.5e-63;
Matches 533; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACCATCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
DB 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTTGAAGGACGGTTAATGCG 131
QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
DB 132 CACTACTTCAAGTGTACAGGAAAGGAGAGGGGCAACCCATTTGAGGGTACGCGAGGAATG 191
QY 127 AAGATCGAGGTGATCGAGGGGGGCGCCCTGCGCTTCGCTTCACATCCTGTCCACCTCC 186
DB 192 AAGATAGAGTTCATCGAGGAGTCCATTGCCATTTGCTTCACATTTGTCAACGAGT 251
QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAAGTACGTGTCGGGATCCCGACTACTTCAAG 246
DB 252 TGTATGTACGGTGTAGAGACCTTCATCAAGTATGTGTGAGGAAATTCCTGACTACTTCAAG 311
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGGAGCGGACCAACCATCGAGGACGGCGGCTTC 306
DB 312 CAGTCTTCCCTGAGGTTTACTTGGGAAAGAACCAACCATCGAGGATGAGGCTTT 371
QY 307 CTGACCGGCCACACGACACCTCCCTCGACGGGACTGCTGTGTGTAACAGGTCAAGATC 366
DB 372 CTTACAGTCTATCAGGACACAGGCTAGATGGAGATTGCCCTGTTTACAGGTCAAGATT 431
QY 367 CTGGGACACACTTCCCGCGGCGCGCGTGTGATGAGCAACAGGCGCGCTGGAG 426
DB 432 CTTGGTAATAATTTCTGCTGTGATGGCCCGTGTGATGAGCAACAGGAGATGGAG 491
QY 427 CCCTCCACCGAGATCGTGTACAGGTGGAGCGGCTGTGCGGCGGCGAGTCCCTGATGGCC 486
DB 492 CCAGGACCGAGATGTTTATGAGTTGACGGTGTGCTGCTGTGACAGACTCTTTGATGGCC 551
QY 487 CTGAGTGCCTCCCGCGGCTGCGACCTGCGACCTGCGACACCACTTACCGCTCCAG 546
DB 552 CTTAAGTGCCCTGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
QY 547 AAGCCGCGCTCCGCGCTGAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
DB 612 AATCAGTAGTGTGCTTGAAGTSCAGGATTTTCAATTTGAAGTATCATCGATCGAGATA 671
QY 607 CTGAGAGGTGAGAGGAGCAAGTGTACAAGCAGTACAGGCGCGGCTGAGGCGCGCTAC 666
DB 672 ATGAGGAAGTTGAGAAAGGCAAGTGTATAAAGTACGAGCAGCAGTGTGCGGAGTAC 731
QY 667 TGCCAGCGCGCTCCAGCTGCGGCGCAACTGA 702
DB 732 TGTGATGCTGCTCCATCCAGCTTGGACATAACTAA 767

RESULT 12
IDA00801 standard; cDNA; 767 BP.
XX
AC ABA00801;
XX
DT 01-APR-2003 (first entry)
XX
DE Kindling fluorescent protein, AsP595 P90L A148G H203Y cDNA.
XX
KW Gene; kindling fluorescent protein; kindling stimulus; movement;
KW labeling; fluorescence resonance energy transfer; FRET;
KW bioluminescence resonance energy transfer; BRET; biosensor;
KW automated screening; ss.
XX
```

```
OS Anthozoa sulcata.
XX
FH Key Location/Qualifiers
XX CDS 59..763
FT /*tag= a
FT /product= "AsP595 P90L A148G H203Y"
FT replace(330,T)
FT /*tag= b
FT replace(496,C)
FT /*tag= c
FT replace(657..659, CAT)
FT /*tag= d
XX
XX WO200296924-A1.
XX
XX 05-DEC-2002.
XX
XX 24-MAY-2002; 2002WO-US016379.
XX
XX 25-MAY-2001; 2001US-0293752P.
XX 11-OCT-2001; 2001US-0329176P.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukanov SA, Chudakov D, Lukyanov K;
XX
XX WPI: 2003-156788/15.
XX P-PSDB; AAG79764.
XX
XX Novel nucleic acid that is present in other than its natural environment
XX and that encodes kindling fluorescent protein, is useful in labeling
XX protocols, e.g. labeling proteins, organelles, cells and organisms.
XX
XX Example; Fig 3; 96pp; English.
XX
XX The sequences given in ABA00799-805 encode wild type and mutant kindling
XX fluorescent proteins. The proteins go from a first substantially non-
XX fluorescent or non-fluorescent state to a second fluorescent state upon
XX exposure to a kindling stimulus. The kindling proteins are useful for
XX detecting an entity such as a protein, organelle or cell in a composition
XX such as a cell or a multicellular composition (preferably a multicellular
XX organism), by providing the entity as an entity labeled with the kindling
XX protein, kindling the kindling fluorescent protein label with a kindling
XX stimulus to produce a kindling fluorescent protein label, and
XX exciting the kindling fluorescent protein label with light and
XX detecting any fluorescence from it to detect the entity. The method
XX monitors the movement of the entity. The fluorescent proteins and the
XX cDNA encoding them are useful in labeling protocols, e.g., labeling
XX proteins, organelles, cells and organisms, as biological labels or
XX markers, in protein labeling or tagging applications. The fluorescent
XX kindling proteins are useful as detectable labels, as labels in analyte
XX detection assays, in fluorescence resonance energy transfer (FRET)
XX applications, in bioluminescence resonance energy transfer (BRET)
XX applications, as biosensors in prokaryotic and eukaryotic cells, in
XX applications involving the automated screening of arrays of cells
XX expressing fluorescent reporting groups, in high throughput screening
XX assays, as second messenger detectors, and in fluorescent activated cell
XX sorting assays
XX
XX Sequence 767 BP; 225 A; 163 C; 188 G; 191 T; 0 U; 0 Other;
```

```
Query Match 61.3%; Score 433.6; DB 7; Length 767;
Best Local Similarity 76.4%; Pred. No. 2.7e-63;
Matches 532; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACCATCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
DB 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTTGAAGGACGGTTAATGCG 131
QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
DB 132 CACTACTTCAAGTGTACAGGAAAGGAGAGGGGCAACCCATTTGAGGGTACCGAGGAATG 191
```

```
QY 127 AAGATCAGAGTGTATCAGAGGCGCGCCCTGCGCTTCCATCCTGTCCACCTCC 186
DB 192 AAGATAGAGTGTATCAGAGGAGTCCATTGCCATTGCTTCCACATTTTGTCAAGAGT 251
QY 187 TGCATGTAGCGTCCAAAGCGCTTCAATCAAGTACGTGTCCGGATCCCGACTACTTCAAG 246
DB 252 TGTATGTAGCGTGTAAAGACCTTCAATCAAGTATGTGTCAAGAAATCTCTGACTACTTCAAG 311
QY 247 CAGTCCCTCCCGAGGCGCTTCACTGGGAGCGCACCACTACGAGGACCGCGCTTC 306
DB 312 CAGTCTTTCCCTGAAGGTCTTACTTGGGAAGACCACTACGAGGATGGAGGCTTT 371
QY 307 CTGACCGCCACAGACACCTCCCTGGAGCGGACTGCTGTGTGTAAGGTGAAGATC 366
DB 372 CTTACAGCTCATCAGGACACAAGCTTAGATGGAGATTGCTGCTTTTACAAGTCAAGATT 431
QY 367 CTGGGCAACAACCTTCCCGCGAGCGCCCGTGTATGCAGACAAGCGCGCGCTGGAG 426
DB 432 CTTGGTAATATTTTCTGCTATGGCCCGGTGATGCAGAAACAAGCAGGAAGATGGAG 491
QY 427 CCCTCCACCGAGATCGTACGAGGTGAGCGGCTGCTCGCGGCGAGTCCCTGATGCC 486
DB 492 CCAGGCACCGAGATGTTTATGAAGTTGACGGTGTCTGCTGGACAGTCTTTGATGGCC 551
QY 487 CTGAGTGTCCCGCGCGTGTGCGACCTGACCTGCCACTGCACACCACTACCGCTCCAG 546
DB 552 CTTAAGTGTCCCTGCTGCTATCTGACTTGCATCTCCATCTCCATCTTACAGGTCCTCAA 611
QY 547 AAGCCGCGCTCCCGCTGAGATGCGCGCTTCCACTTCGAGGACCAACCGATCGAGATC 606
DB 612 AAACAGCTAGTGCCTTGAAGTGCAGGATTTTATTTGAAGATTACCGCATCGAGATA 671
QY 607 CTGAGGAGTGTGAGAGGCAAGTGTCTCAAGCAGTACGAGGCGCGCTGGCGCGCTAC 666
DB 672 ATGGAGGAGTGTGAGAGGCAAGTGTCTATAACAGTACGAGCAGTGGCGCGGTAC 731
QY 667 TGGAGCGCGCCCTCCAGCTGGCGCAACTGA 702
DB 732 TGTATGCTGCTCTCATCAAGCTTGGACATACTAA 767

RESULT 13
ABL41173
ID ABL41173 standard; cDNA; 687 BP.
XX AC ABL41173;
XX
XX 12-AUG-2002 (first entry)
XX
XX H. crispa alternative fluorescent protein mutant FP10-cr1 cDNA.
XX
XX Stichodactyliden; chromoprotein; fluorescent; anthozoan; food; FRET;
KW colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;
KW fluorescence resonance energy transfer; gene expression; mutant; ss.
XX
XX Heteractis crispa.
XX
XX Key Location/Qualifiers
XX CDS 1..687
XX FT /*tag= a
XX FT /product= "mutant fluorescent protein"
XX misc_feature 4..6
XX FT /*tag= b
XX FT /note= "the amino acid Val encoded by the above codon is
XX not indicated in the corresponding protein"
XX
XX WO200230965-A2.
XX
XX 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US032080.
XX
XX 12-OCT-2000; 2000US-0240018P.
XX PR
```

```
PR 16-JUL-2001; 2001US-0306131P.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukyanov SA, Fradkov AP, Lukyanov KA, Gurskaya NG;
XX
XX WPI; 2002-444170/47.
XX P-PSDB; ABB07995.
XX
XX Novel nucleic acid encoding Stichodactyliden chromoprotein and its
XX fluorescent mutant useful as coloring agent, labels in analyte detection
XX assays, markers in recombinant DNA applications and filters in
XX sunscreens.
XX
XX Claim 5; Fig 10; 81pp; English.
XX
XX The invention relates to a nucleic acid present in other than its natural
XX environment and encoding an Stichodactyliden chromoprotein or its
XX fluorescent mutant, where the fluorescent protein has an emission maximum
XX ranging from 580-660 nm. the polynucleotides and encoded proteins are
XX useful in applications employing a chromo or fluorescent nucleic acid or
XX protein. Recombinant vectors comprising the nucleic acid is useful for
XX producing an Anthozoan chromo and/or fluorescent protein. The
XX chromoproteins, and their fluorescent mutants are useful as colouring
XX agents capable of imparting colour or pigment to a particular composition
XX of matter. The chromoproteins can be incorporated into a variety of
XX different compositions including food compositions, pharmaceuticals, in
XX cosmetics, living organisms, e.g. animals and plants, and as labels in
XX analyte detection assays, e.g. assays for biological analyses of interest
XX (see ABL41167 for a detailed description of the various uses of the
XX chromoproteins). The present sequence represents the H. crispa
XX alternative fluorescent protein mutant FP10-cr1 encoding cDNA
XX
XX Sequence 687 BP; 145 A; 230 C; 219 G; 93 T; 0 U; 0 Other;
XX
XX Query Match 59.1%; Score 417.8; DB 6; Length 687;
XX Best Local Similarity 79.1%; Pred. No. 1.1e-60;
XX Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;
XX
XX QY 12 CTTGCTGACCGAGACCATGCCCTTCAGGACCACCATCGAGGCGACCTGTGACGCCACTA 71
XX DB 12 CTTGCTGAGGAGAGCATGCGCATCAAGATGTATGAGGCGACCTGTGACGCCACTA 71
XX
XX QY 72 CTTCAAGTGCACCGGCGAGGCGCAACCCCTCGAGGCGACCCAGGAGATGAAGAT 131
XX DB 72 CTTCAAGTGCAGGCGAGGCGGACGCAACCCCTCGCGCGCACCCAGAGATGCGAT 131
XX
XX QY 132 CGAGTGTATCAGGCGGCGCCCTGCGCTTCCATCTCCATCTGCTCCACTCTGCTGAT 191
XX DB 132 CCACGTGACCGAGGCGCGCCCTGCGCTTCCATCTGCTCCACTCTGCTGCTGCGA 191
XX
XX QY 192 GTAGCGCTCCAAAGCGCTTCAATCAAGTACGTGTCCGGCATCCCGACTACTTCAAGCAGTC 251
XX DB 192 GTAGCGGAGGAGGACCTTGTGACCAACACCGCGAGATCCCGACTTCTTCAAGCAGAG 251
XX
XX QY 252 CTTCCCGAGGCGCTTCACTCGGAGCGCACCAACCTTACGAGACGCGGCTTCTGAC 311
XX DB 252 CTTCCCGAGGCGCTTCACTCGGAGAGAAACCAACCTTACGAGACGCGGCTTCTGAC 311
XX
XX QY 312 CGCCACCGAGGACCTTCCCTGAGCGGCTGCTGCTGTGTACAGGTGAGATCTCTGGG 371
XX DB 312 CGCCACCGAGGACCTTCCCTGAGGCGGCACTGCTGATCTTACAGGTGAGATCTCTGGG 371
XX
XX QY 372 CAACAATTTCCCGCGAGCGCCCTGATGACAGACAAGCGCGCGCTCGGAGCCCTC 431
XX DB 372 CACCAACTTCCCGCGAGCGCCCTGATGATGAAGACAAGAGCGCGGCTGGAGCCCGAG 431
XX
XX QY 432 CACCGAGATCGTGTACAGGTTGAGCGGCTGCTGCGGCGGCGAGTGTGAGGCTTGA 491
XX DB 432 CACCGAGGTTGTATCCCGAGGAGCGGCTGCTGCGGCGGCGGAAACGTGTGCGGCTTGA 491
XX
XX QY 492 GTGCGCGCGGCTGCGGACCTGACCTGACCTGACACCACTACCGCTCCAAAGAGCC 551
XX DB 492 GTGCGCGCGGCTGCGGACCTGACCTGACCTGACACCACTACCGCTCCAAAGAGCC 551
```


PH	Key	Location/Qualifiers
FT	CDS	26. .472
FT		/*tag= a
FT		/product= "asCP562"
XX		
	DE10026520-A1.	
	PN	

Blank Sheet as PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:23:50 ; Search time 50.2812 Seconds
(without alignments)
7945.752 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300
Sequence: 1 ggcacgcctccctgctgac.....ctgggccacaactgaagctt 707

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ntp.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_29072004_150505_25453/app.query.fasta_1.1742
-DB=A Geneseq 29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10081864@cgn_1_128 @runat_29072004_150505_25453 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 29Jan04:*

1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1257	96.7	231	5 AAE28841	Aae28841 Anemonia
2	1257	96.7	466	6 ABU09925	Abu09925 Chromopro
3	1256	96.5	468	6 ABU09924	Abu09924 Chromopro
4	1251	96.2	232	5 AAE28842	Aae28842 Anemonia
5	1226	94.3	231	3 AAY99734	Aay99734 Anemonia
6	1223	94.1	232	4 AAE00376	Aae00376 Anemonia
7	1220	93.8	232	4 AAE00385	Aae00385 Anemonia
8	1219	93.8	232	3 AAY99733	Aay99733 Anemonia
9	1219	93.8	232	3 AAB01623	Aab01623 Anemonia
10	1219	93.8	232	5 ABP70043	Abp70043 Colour Fa

11	1219	93.8	232	6 AAG79762	Aag79762 Wild type
12	1218	93.7	232	6 AAG79763	Aag79763 Kindling
13	1216	93.5	232	5 AAE28834	Aae28834 Anemonia
14	1206	92.8	232	6 AAG79764	Aag79764 Kindling
15	786.5	60.1	226	5 AAG79467	C. gigant
16	781.5	60.1	227	5 AAG79466	C. gigant
17	780.5	60.0	227	6 AAG79769	H. crispa
18	771.5	59.3	458	6 ABU09923	Abu09923 Chromopro
19	770.5	59.3	226	5 ABB07993	H. crispa
20	770.5	59.3	227	6 ABB07968	H. crispa
21	769.5	59.2	226	5 ABB07992	H. crispa
22	769.5	59.2	460	5 ABB07996	Cr-449-ta
23	769.5	59.2	460	5 ABU09921	Chromopro
24	769.5	59.2	470	5 ABB07997	Cr-449-ta
25	769.5	59.2	470	5 ABU09922	Chromopro
26	768.5	59.1	227	5 ABB07998	H. crispa
27	767.5	59.0	227	5 ABB07991	H. crispa
28	766.5	59.0	226	5 ABB07999	H. crispa
29	766.5	59.0	227	5 ABB07994	H. crispa
30	763.5	58.7	227	6 AAG79765	H. crispa
31	762.5	58.7	227	5 ABB07995	H. crispa
32	762.5	58.7	227	5 ABB07989	H. crispa
33	762.5	58.7	227	5 ABB08000	H. crispa
34	757.5	58.3	227	6 AAG79766	H. crispa
35	755.5	58.2	227	6 AAG79767	H. crispa
36	753.5	58.0	227	5 ABB07990	H. crispa
37	584	44.9	228	5 ABG67284	Polythoa
38	583	44.8	228	5 ABG67283	Polythoa
39	583	44.8	228	5 ABG67278	Polythoa
40	583	44.8	228	5 ABG67282	Polythoa
41	583	44.8	228	5 ABG67281	Polythoa
42	583	44.8	228	5 ABG67285	Consensus
43	583	44.8	228	5 ABG67277	Polythoa
44	583	44.8	303	5 ABG67279	Polythoa
45	575.5	44.3	229	6 AAO29587	Condyllact

ALIGNMENTS

RESULT 1

AAE28841
ID AAE28841 standard; protein; 231 AA.

XX
AC AAE28841;
XX

DT 27-DEC-2002 (first entry)

XX Anemonia sulcata asFP595 non-aggregating mutant protein, FP7-NA.

XX Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; asFP595 protein; mutant; mutein.

XX Anemonia sulcata.

OS Synthetic.

XX WO200268459-A2.

XX 06-SEP-2002.

PD 20-FEB-2002; 2002WO-US005749.

XX 21-FEB-2001; 2001US-0270983P.

PR 04-DEC-2001; 2001US-00006922.

XX (CLON-) CLONTECH LAB INC.

PA Lukanov S, Lukanov K, Yanushevich Y, Savitsky A, Fradkov A;

PI WPI; 2002-691654/74.

XX N-PSDB; AAD46286.

XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.

XX Disclosure; Page 77; 80pp; English.

XX The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Anemonia sulcata asFP595 non-aggregating mutant protein of
 CC the invention

XX SQ Sequence 231 AA;

Alignment Scores:
 Pred. No.: 7,82e-90 Length: 231
 Score: 1257.00 Matches: 231
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.69% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x AAE28841 (1-231)

QY 7 GCCTCCCTGTCAGCCGAGACCATCCCTTCAGACCAACCATCGAGGCGACCGTGAACGGC 66
 DB 1 AlaSerLeuThrGluThrMetProPheArgThrIleGluGlyThrValAsnGly 20
 QY 67 CACTACTTCAAGTCACCGGAGGGGAGGGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
 DB 21 HistyrPheLysCysThrGlyLysGlyGluGlyLysProLeuGluGlyThrGlnGluMet 40
 QY 127 AAGATCGAGGTGATCGAGGGGGCCCTCCCTTCGCTTCACATCTGTCCTCCACTCC 186
 DB 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
 QY 187 TGCATGTACGGCTCCAGGGCTTCATCAAGTACGTGTCCGGCATCCCGACTCTCTCAG 246
 DB 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
 QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGGAGCGGACCAACCACTACGAGGAGCGGCGCTTC 306
 DB 81 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTrpGluAspGlyGlyPhe 100
 QY 307 CTGACCGGCCACAGGACACTCCCTGAGGGGAGCTGCTGTGTACAGGTGAAGATC 366
 DB 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
 QY 367 CTGGGCAACAACTTCCCGCGAGCGGCCGCTGTATGCAGAACAAAGCGCGCTCGGAG 426
 DB 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyTrpGlu 140
 QY 427 CCCTCCACCGAGATCGTGTGTCAGAGTGGACGGCGTGTGCGCGCCAGTCCCTGTATGCC 486
 DB 141 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
 QY 487 CTGAGGTGCCCGCGCTCCGCTGACCTGACCTGCACCTGCACACCACTACCGCTCCAG 546
 DB 161 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 180
 QY 547 AAGCCCGCTCCGCGCTTGAAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
 DB 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 200
 QY 607 CTGGAGGAGTGGAGAGGCGAGTGTCTACAGCATGACGAGCGCGCTGGCGCGCTAC 666

DB 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 220
 QY 667 TCGAGCGCGCGCCCTCCCAAGCTGGCCACAC 699
 DB 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231

RESULT 2

ABU09925
 ID ABU09925 standard; protein; 466 AA.
 XX AC ABU09925;
 XX DT 09-JUL-2003 (first entry)
 XX DE Chromoprotein AsRed-35-5D-tandem.
 XX KW Chromo/fluorescent domain; labeled fusion protein;
 XX KW site-specific gene modification; chromoprotein; colouring agent;
 XX KW food composition; pharmaceutical; cosmetic; AsRed-35-5D-tandem.
 XX OS Anthozoa.
 XX PN WO2003031590-A2.
 XX PD 17-APR-2003.
 XX PF 10-OCT-2002; 2002WO-US032560.
 XX PR 12-OCT-2001; 2001US-00976673.
 XX PR 11-FEB-2002; 2002US-0356225P.
 XX PR 22-MAY-2002; 2002US-0383338P.
 XX PA (CLON-) CLONTECH LAB INC.
 XX PI Lukanov SA;
 XX DR MPI: 2003-381709/36.
 XX DR N-PSDB; ACA61028.
 XX PT New nucleic acid encoding polypeptide products having at least two linked
 PT chromo/fluorescent domains, useful for generating transgenic plants or
 PT animals or site-specific gene modifications in cell lines.
 PS Disclosure; Fig 5; 68pp; English.

CC The invention describes a nucleic acid encoding a polypeptide product
 CC comprising a first and a second chromo/fluorescent domain, optionally
 CC joined by a linking domain. The first and second chromo/fluorescent
 CC domains associate with each other under intracellular conditions so that
 CC the encoded polypeptide assumes a tertiary structure. The nucleic acid
 CC and the protein are useful in producing labeled fusion proteins that have
 CC a precise and predictable signal to fusion partner ratio. The nucleic
 CC acid may also be used in generating transgenic, non-human plants or
 CC animals or site-specific gene modifications in cell lines. The
 CC chromoproteins may be used as colouring agents, as a food composition, in
 CC pharmaceuticals or cosmetics, as labels in analyte detection assays or as
 CC selectable markers in recombinant DNA applications. This is the amino
 CC acid sequence of chromo/fluorescent domain fusion protein AsRed-35-5D-
 CC tandem

SQ Sequence 466 AA;

Alignment Scores:
 Pred. No.: 9.11e-90 Length: 466
 Score: 1257.00 Matches: 231
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.69% Indels: 0
 DB: 6 Gaps: 0

US-10-081-864-21 (1-707) x ABU09925 (1-466)

QY 7 GCCTCCCTGTCAGCGAGACCATCCCTTCAGGACCACTCGAGGCGACCGTGAACGGC 66

Db 1 AlaSerLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTCAAGTGCACCGGAGGCGGAGGCAACCCCTCGAGGGGACCCAGAGATG 126
Db 21 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 40
QY 127 AGAGTCAGGTGATCAGAGGCGGCGCCCTCGCTTCGCTTCCATCCTGTCACCTCC 186
Db 41 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATCTACGGCTCCAAAGCGCTTCATCAAGTACGTGTCGGGCATCCCGACTACTCAAG 246
Db 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCTCCCGAGGCTTCACCTGGAGCGGCGGACCCACCTACGAGGACCGGCTTC 306
Db 81 GlnSerLeuProGluGlyPheThrTrpGluA-gThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACGCGCCACGAGACACCTCCCTGGACGGCGACTGCTGTGTACAGGTGAAGATC 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValIle 120
QY 367 CTGGGCAACAACTTCCCGCGGAGCGGCGGCTGATGAGAACAGGCGCGCTGGGAG 426
Db 121 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 140
QY 427 CCTCTCACGAGATCTGTACGAGGTGGAGCGGCTGCTGGCGGCCAGTCCCTGATGGCC 486
Db 141 ProSerThrGluIleValIleGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGAGTGGCCCGGCGGTCCCACTGACCTGCCACTGCACACCACTACCGCTCCAAG 546
Db 161 LeuGluCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTrpArgSerLys 180
QY 547 AAGCCGCTCCCGCTGAAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 200
QY 607 CTGAGGAGGTGGAGAGGCAAGTCTACAAGCAGTACGAGCGCGCTGGCGCGCTAC 666
Db 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 220
QY 667 TGGAGCGCCCGCTCCAAAGCTGGGCGCAAC 699
Db 221 CysAspAlaProSerLysLeuGlyHisAsn 231
RESULT 3
ABU09924
ID ABU09924 standard; protein; 468 AA.
XX AC ABU09924;
XX XX
DT 09-JUL-2003 (first entry)
XX DE Chromoprotein AsRed-35-SNA-tandem.
XX KW Chromo/fluorescent domain; labeled fusion protein;
KW site-specific gene modification; chromoprotein; colouring agent;
KW food composition; pharmaceutical; cosmetic; AsRed-35-SNA-tandem.
XX OS Anthozoa.
XX FN WO2003031590-A2.
XX PD 17-APR-2003.
XX XX
PP 10-OCT-2002; 2002WO-US032560.
XX PR 12-OCT-2001; 2001US-00976673.
XX PR 11-FEB-2002; 2002US-0356225P.
XX PR 22-MAY-2002; 2002US-0383336P.
XX XX

(CLON-) CLONTECH LAB INC.

Lukyanov SA;

WPI; 2003-381709/36.

N-PSDB; ACA61027.

New nucleic acid encoding polypeptide products having at least two linked chromo/fluorescent domains, useful for generating transgenic plants or animals or site-specific gene modifications in cell lines.

Disclosure; Fig 4; 68pp; English.

The invention describes a nucleic acid encoding a polypeptide product comprising a first and a second chromo/fluorescent domain, optionally joined by a linking domain. The first and second chromo/fluorescent domains associate with each other under intracellular conditions so that the encoded polypeptide assumes a tertiary structure. The nucleic acid and the protein are useful in producing labeled fusion proteins that have a precise and predictable signal to fusion partner ratio. The nucleic acid may also be used in generating transgenic, non-human plants or animals or site-specific gene modifications in cell lines. The chromoproteins may be used as colouring agents, as a food composition, in pharmaceuticals or cosmetics, as labels in analyte detection assays or as selectable markers in recombinant DNA applications. This is the amino acid sequence of chromo/fluorescent domain fusion protein AsRed-35-SNA-tandem

Sequence 468 AA;

Alignment Scores:

Pred. No.: 1,09e-89 Length: 468
Score: 1256.00 Matches: 231
Percent Similarity: 99.14% Conservative: 0
Best Local Similarity: 99.14% Mismatches: 2
Query Match: 96.62% Indels: 0
DB: Gaps: 0

US-10-081-864-21 (1-707) x ABU09924 (1-468)

QY 1 GGCATCCGCTCCCTCTCAACGAGACCATGCCCTTCAGACACCATCGAGGGCACCGTG 60
Db 236 GlyMetAlaSerLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrVal 255
QY 61 AACGCCCACTACTTCAAGTGCACCGGCAAGGGGAGGGGCAACCCCTCGAGGGCACCCAG 120
Db 256 AsnGlyHisTyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGln 275
QY 121 GAGATGAAGATCGAGGTGATCGAGGGCGGCCCTTCGCCCTTCCACATCCTGTCC 180
Db 276 GluMetLysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSer 295
QY 181 ACCTCCTGCATGACGGCTCAAGCCCTTCATCAAGTACGTGTCGGGATCCCGACTAC 240
Db 296 ThrSerCysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyr 315
QY 241 TTCAGAGCATCCCTCCCGAGGGCTTCACCTGGAGCGCACCCACCATCGAGGACGGC 300
Db 316 PheLysGlnSerLeuProGluGlyPheThrTrpGluA-gThrThrTyrGluAspGly 335
QY 301 GGCCTTCCTGACCGCCCAACGAGACACCTCCCTCGAGCGGCGACTGCTGGTGTACAGGTG 360
Db 336 GlyPheLeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysVal 355
QY 361 AAGATCCTGGGCAACAACCTTCCCGCGGCGGCGGCGGTGATCGAGAACAGGCGCGCGC 420
Db 356 LysIleLeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArg 375
QY 421 TGGGAGCCCTCCACCGAGATCCTGTACGAGGTGGAGCGGCTGCTCGGCGCGAGTCCCTG 480
Db 376 TrpGluProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerSer 395
QY 481 ATGGCCCTGGAGTGCCCGCGGCTGCGCACCTGACCTGCCACCTGCACACCACTACCGC 540

PD 15-JUN-2000.
 XX
 PF 10-DEC-1999; 99WO-US029300.
 XX
 PR 11-DEC-1998; 98US-00210330.
 PR 09-DEC-1999; 99US-00210330.
 XX
 PA (CLON-) CLONTECH LAB INC.
 XX
 PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Jiang X, Duong T;
 PI Zhao X;
 XX
 DR WPI; 2000-423374/36.
 DR N-PSDB; AA50884.
 XX
 PT Novel fluorescent proteins from non-bioluminescent *Anemonia sulcata*,
 PT useful for fluorescent labeling and as markers.
 XX
 PS Claim 9; Page 66-67; 71pp; English.
 XX
 CC The present sequence is the mutant fluorescent protein, Mut1, from non-
 CC bioluminescent *Anemonia sulcata*. The fluorescent protein has applications
 CC in fluorescent labelling, as fluorescent markers for gene expression and
 CC protein localisation studies, and in fluorescence resonance energy
 CC transfer (FRET) reactions. The coding sequence may be used as a source or
 CC primers and probes for identifying related proteins
 XX
 SQ Sequence 231 AA;

Alignment Scores:
 Pred. No.: 2,07e-87 Length: 231
 Score: 1226.00 Matches: 223
 Percent Similarity: 98.27% Conservative: 4
 Best Local Similarity: 96.54% Mismatches: 4
 Query Match: 94.31% Indels: 0
 DB: Gaps: 0

US-10-081-864-21 (1-707) x AAY99734 (1-231)

QY 7 GCTCCCTCTGACCGAGACCATGCCCTTCAGGACACCATCGAGGACACCGTGAACGCG 66
 Db 1 AlaSerPheLeuLysThrMetProPheLysThrIleGluGlyThrValAsnGly 20
 QY 67 CACTACTCAAGTGCAACCGGCAAGGCGGAGGCGCAACCCCTCGAGGCGCACCCAGAGATG 126
 Db 21 HisTyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
 QY 127 AAGATCGAGGTATCGAGGCGGCGCCCTCGCCCTTCGCTTCACATCCGTGTCACCTCC 186
 Db 41 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
 QY 187 TGCAATAGCGCTCCCAAGGCTTCATCAAGTAGCTGTCGCGCATCCCGACTACTTCAAG 246
 Db 61 CysMetTyGlySerLysAlaPheIleLysTyValSerGlyIleProAspTyPheLys 80
 QY 247 CAGTCCTCCCGAGGGCTTACCTGGAGCGGCACCCACCATCGAGGACGCGGCTTC 306
 Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyTrpGluAspGlyGlyPhe 100
 QY 307 CTGACCGCCACCGAGACACCTCCCTGGACGGGACTGCTGTGTACAAAGTGAAGATC 366
 Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyLysValIle 120
 QY 367 CTGGGCAACAACATCCCGCGCGCGCCCGTGCATGCAGAACAGCGCGCGCTGGGAG 426
 Db 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 140
 QY 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGCTGCGGCGCCAGTCCCTGATGGCC 486
 Db 141 ProSerThrGluIleValTyTrpGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
 QY 487 CTGGAGTGCCCGCGGTCCCACTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 546

Db 161 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyArgSerLys 180
 QY 547 AAGCCGCGCTCCGCGCTCGAAGATGCCGCTTCACCTTCGAGGACACCGCATCGAGATC 606
 Db 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheHisGluAspHisArgIleGluIle 200
 QY 607 CTGAGGAGGTGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCTGGCGCGCTAC 666
 Db 201 MetGluGluValGluLysGlyLysCysTyLysGlnTyTrpGluAlaValGlyArgTy 220
 QY 667 TGGAGCGCGCGCGCTCCCAAGCTGGGCGCAAC 699
 Db 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231

RESULT 6
 AAE00376
 ID AAE00376 standard; protein; 232 AA.
 XX
 AC AAE00376;
 XX
 XX 19-JUN-2001 (first entry)
 DT
 XX
 DE *Anemonia sulcata* chromo/fluorescent protein, asFP600 (NFP-7).
 XX
 KW Anthozoa; Chromoprotein; fluorescent protein; asFP600; NFP-7; sunscreen;
 KW analyte detection assay; selectable marker; recombinant DNA application;
 KW biosensor; pH indicator; invivo marker; selective filter; sea anemone.
 XX
 OS *Anemonia sulcata*.
 XX
 FN WO200127150-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 13-OCT-2000; 2000WO-US028477.
 XX
 PR 14-OCT-1999; 99US-00418529.
 PR 15-OCT-1999; 99US-00418917.
 PR 15-OCT-1999; 99US-00418922.
 PR 19-NOV-1999; 99US-0044338.
 PR 19-NOV-1999; 99US-0044341.
 PR 09-DEC-1999; 99US-00457556.
 PR 09-DEC-1999; 99US-00457898.
 PR 09-DEC-1999; 99US-00458144.
 PR 09-DEC-1999; 99US-00458477.
 PR 10-DEC-1999; 99WO-US029405.
 PR 14-JUN-2000; 2000US-0211607P.
 PR 14-JUN-2000; 2000US-0211609P.
 PR 14-JUN-2000; 2000US-0211626P.
 PR 14-JUN-2000; 2000US-0211627P.
 PR 14-JUN-2000; 2000US-0211687P.
 PR 14-JUN-2000; 2000US-0211766P.
 PR 14-JUN-2000; 2000US-0211880P.
 PR 14-JUN-2000; 2000US-0211888P.
 PR 14-JUN-2000; 2000US-0212070P.
 XX
 XX (CLON-) CLONTECH LAB INC.
 PA
 XX
 PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Tersikh A;
 XX
 XX WPI; 2001-266409/27.
 XX N-PSDB; AAD03616.
 DR
 DR
 XX
 PT An Anthozoa chromo- or fluorescent protein (PI) present in an environment
 PT other than its natural environment, useful as a label in analyte/
 PT detection assays and as a selectable marker in recombinant DNA
 PT applications.
 XX
 PS Claim 8; Fig 7; 69pp; English.
 CC
 CC The present sequence is an *Anemonia sulcata* chromo/fluorescent protein,
 CC asFP600 (NFP-7). NFP-7 is present in an environment other than its
 CC natural environment and has an absorbance maximum ranging from 370nm to

CC 770nm and more usually from 560nm to 580nm and emission maximum ranging
 CC from 395nm to 795nm and more usually from 585 to 605nm. The
 CC chromoproteins or fluorescent proteins are useful as labels in analyte
 CC detection assays, as selectable markers in recombinant DNA applications,
 CC as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
 CC and as in vivo markers in animals. They are also useful in sunscreens and
 CC as selective filters. Note: The specification also refers to a mutant of
 CC this sequence, Mut1 with substitutions T70A and A148S. However, A at
 CC position 148 do not match with this wild type sequence
 XX
 SQ Sequence 232 AA;

Alignment Scores:
 Pred. No.: 3,56e-87 Length: 232
 Score: 1223.00 Matches: 222
 Percent Similarity: 98.27% Conservative: 5
 Best Local Similarity: 96.10% Mismatches: 4
 Query Match: 94.08% Indels: 0
 DB: 4 Gaps: 0

US-10-081-864-21 (1-707) x AAE00385 (11-232)

QY 7 GCCTCCCTGTCAGCCGAGACATGCTCCCTTCAGGACACCATCGAGGCGACCGTGAACGGC 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTCACCGCAGGGGAGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
 Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCAGGTGATCGAGGCGGCGCCCTCCCTTCCTCCATCTGTCACATCTGTCACCTCC 186
 Db 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAAGTGTCTCGGCGATCCCGCTACTTCAG 246
 Db 62 CysMetTyrGlySerLysAlaPheileLysTyrValSerGlyProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGCGCACCCACCATCTACGAGGCGCGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACCGAGACACTCTCTGAGCGGCGACTGCTGTGTACAGGTGATGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysile 121
 QY 367 CTGGCAACCACTTCCCGCGAGCGCCCGTGTATGAGAACAGCGCGCGCTGGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTTCACCGAGATCGTGTACGAGGTGAGCGCGTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 142 ProSerThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTCCCGCGCTGCGACCTGACCTGCGACCTGACACACCTACCGCTCCAG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTGAGATGCGCGCTTCACCTTCGAGGACCGACCGATCGATC 606
 Db 182 LysProAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgileGluile 201
 QY 607 CTGGAGGTGAGGAGGCGAGTCTACAGCAGTACAGGCGCGCGCGCGCTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGAGCGCGCGCGCTCCAGTGGGCCCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 7
 AAE00385
 ID AAE00385 standard; protein; 232 AA.

XX AAE00385;
 AC 19-JUN-2001 (first entry)
 XX
 DT Anemonia sulcata asFP600 (NFP-7) mutant (A184S).
 XX
 DE Anthozoa; Chromoprotein; fluorescent protein; asFP600; NFP-7; sunscreen;
 KW analyte detection assay; selectable marker; recombinant DNA application;
 KW biosensor; pH indicator; invivo marker; selective filter; sea anemone;
 KW mutant; mutein.
 XX
 OS Anemonia sulcata.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 184
 FT /note= "Wild type Ala substituted with Ser"
 FT
 XX WO200127150-A2.
 XX
 PD 19-APR-2001.
 XX
 XX 13-OCT-2000; 2000WO-US028477.
 XX
 PR 14-OCT-1999; 99US-00418529.
 PR 15-OCT-1999; 99US-00418917.
 PR 15-OCT-1999; 99US-00418922.
 PR 19-NOV-1999; 99US-00444338.
 PR 19-NOV-1999; 99US-00444341.
 PR 09-DEC-1999; 99US-00457556.
 PR 09-DEC-1999; 99US-00457898.
 PR 09-DEC-1999; 99US-00458144.
 PR 09-DEC-1999; 99US-00458477.
 PR 10-DEC-1999; 99WO-US029405.
 PR 14-JUN-2000; 2000US-0211607P.
 PR 14-JUN-2000; 2000US-0211609P.
 PR 14-JUN-2000; 2000US-0211626P.
 PR 14-JUN-2000; 2000US-0211627P.
 PR 14-JUN-2000; 2000US-0211687P.
 PR 14-JUN-2000; 2000US-0211766P.
 PR 14-JUN-2000; 2000US-0211880P.
 PR 14-JUN-2000; 2000US-0211888P.
 PR 14-JUN-2000; 2000US-0212070P.
 XX
 PA (CLON-) CLONTECH LAB INC.
 XX
 PI Lukanov SA, Fradkov AF, Labas YA, Matz MV, Tersikh A;
 XX
 DR WPI; 2001-265409/27.
 XX
 PT An Anthozoa chromo- or fluorescent protein (PI) present in an environment
 PT other than its natural environment, useful as a label in analyte
 PT detection assays and as a selectable marker in recombinant DNA
 PT applications.
 XX
 PS Disclosure; Page: 69pp; English.
 XX
 CC The present sequence is an Anemonia sulcata chromo/fluorescent protein,
 CC asFP600 (NFP-7) mutant (A184S). NFP-7 is present in an environment other
 CC than its natural environment and has an absorbance maximum ranging from
 CC 370nm to 770nm and more usually from 560nm to 580nm and emission maximum
 CC ranging from 395nm to 795nm and more usually from 585 to 605nm. The
 CC chromoproteins or fluorescent proteins are useful as labels in analyte
 CC detection assays, as selectable markers in recombinant DNA applications,
 CC as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
 CC and as in vivo markers in animals. They are also useful in sunscreens and
 CC as selective filters. Note: The present sequence is not shown in the
 CC specification, but derived from asFP600 (NFP-7) sequence (AAE00376) shown
 CC in figure 7
 XX
 SQ Sequence 232 AA;

PN	WO200034319-A1.
XX	
PD	15-JUN-2000.
XX	
PF	10-DEC-1999; 99WO-US029300.
XX	
PR	11-DEC-1998; 98US-00210330.
PR	09-DEC-1999; 99US-00210330.
XX	(CLON-) CLONTECH LAB INC.
PA	
XX	Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Jiang X, Duong T;
PI	Zhao X;
XX	
XX	WPI; 2000-423374/36.
DR	N-PSDB; AAA50863.
XX	
XX	Novel fluorescent proteins from non-bioluminescent Anemonia sulcata,
PT	useful for fluorescent labeling and as markers.
XX	
PS	Claim 9; Page 64-65; 71pp; English.
XX	
CC	The present sequence is the fluorescent protein, asFP600, from non-
CC	bioluminescent Anemonia sulcata. The fluorescent protein has application and
CC	in fluorescent labelling, as fluorescent markers for gene expression and
CC	protein localisation studies, and in fluorescence resonance energy
CC	transfer (FRET) reactions. The coding sequence (AAA50863) may be used as
CC	a source or primers and probes for identifying related proteins
XX	
SQ	Sequence 232 AA;
Alignment Scores:	
Pred. No.:	7,31e-87
Score:	1219.00
Percent Similarity:	97.84%
Best Local Similarity:	95.67%
Query Match:	93.77%
DB:	3
US-10-081-864-21 (1-707) x AA999733 (1-232)	
QY	7 GCCTCCCTGCTGACGAGACCATGCGCTTCAGGACCAACATCGAGGCGACCGTGAACGGC 66
DB	2 AlaSerPheLeuLysLysThrMetPhePheLysThrThrLeuGluGlyThrValAsnGly 21
QY	67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGCAACCCCTCGAGGCGACCCAGGAGATG 126
DB	22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY	127 AAGATCGAGGTGATCGAGGCGGCGCCCTCGCTTCGCTTCACATCCTGTCACCTCC 186
DB	42 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY	187 TGCATGTACGCTCCAAAGGCTTCATCAAGTACGTGTCGGCATCCCGACTACTTCAAG 246
DB	62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY	247 CAGTCCCTCCCGGAGGCTTCACCTGGGAGCGCACCAACCTACGAGGCGCGGCTTC 306
DB	82 GlnSerPheProGluGlyPheThrTyrGluArgThrThrTyrGluAspGlyGlyPhe 101
QY	307 CTGACCGCCACACGAGACACCTCCCTCGAGCGGCGACTGCTGTGTACAGGTGAAGTC 366
DB	102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY	367 CTGGGCAACAACTTCCCGCGGCGGCGCTGATCAGAACAAAGCGCGGCGCTGGGAG 426
DB	122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrGlu 141
QY	427 CCCTCCACCGAGATGTTACAGAGTGGACGCGGCTGCTCGGCGGCGCTCCCTGATGCGC 486
DB	142 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY	487 CTGAGTGCCTCCCGGCTGCGCACCTGACCTGCACCTGCACACACCTACCGCTCCAAG 546
DB	162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY	547 AAGCCGCTCCCGCTGAGATGCGCGGCTTCACCTTCGAGGACCAACCGCATCGAGTC 606
DB	182 LysProSerAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY	607 CTGAGAGGTGGAGAGGCGAGTGCTACAGCAGTACGAGCGCGCGCTGGCGCGCTAC 666
DB	202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
QY	667 TGCGAGCGCGCGCTCCAAAGCTGGGCGCAAC 699
DB	222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232
RESULT 8	
ID	AA999733
XX	AA999733 standard; protein; 232 AA.
XX	
AC	AA999733;
XX	
DT	08-SEP-2000 (first entry)
XX	Anemonia sulcata fluorescent protein, asFP600.
XX	
XX	Green fluorescent protein; GFP; asFP600; protein localisation;
KW	fluorescence resonance energy transfer.
XX	
OS	Anemonia sulcata.
XX	

XX 01-MAR-2002; 2002WO-GB000928.
 XX 02-MAR-2001; 2001US-0273227P.
 PR 21-MAR-2001; 2001AU-00003874.
 PR 15-OCT-2001; 2001US-0329816P.
 XX (NUFA-) NUFARM LTD.
 PA (UYOU) UNIV QUEENSLAND.
 PA (JONE/) JONES E L.
 XX Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
 PI Hoehn-Guldberg IO, Prescott M;
 PI WPI; 2002-740765/80.
 XX Novel color-facilitating molecule for producing a biomatrix, has a
 PT polypeptide which alone/along with molecules imparts altered visual
 PT characteristics to cells in the absence of excitation by extraneous non-
 PT white light.
 XX Example 20; Page 503-504; 510pp; English.
 XX The invention relates to an isolated colour-facilitating molecule (CFM)
 CC comprising a polypeptide which, in a cell, alone or together with one or
 CC more other molecules imparts an altered visual characteristic to the cell
 CC when visualised by a human eye in the absence of excitation by extraneous
 CC non-white light or particle emission. CFMs are useful for producing a
 CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
 CC red coloured fleece. They are useful for producing coloured plant
 CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
 CC uses include transducing or intensifying an image, providing additional
 CC light for growing phototropic organisms e.g. algae and/or corals, for
 CC coating materials that experience UV damage e.g. plastics and car
 CC upholstery. CFMs are useful in the flower industry, in the development of
 CC new varieties of flowering plants. Other contemplated uses include,
 CC expression markers, general reporter molecules, photon traps, UV sinks or
 CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
 CC fungal species, and in fruits and vegetables to enhance their
 CC marketability. CFMs embedded in a gel matrix improve image quality in
 CC situations of distorted light spectra (biomatrix). The first all-protein
 CC chromophore to be isolated was Green Fluorescent protein (GFP). The
 CC sequences given in records ABP69924-ABP70048 represent CFM related amino
 CC acid sequences
 XX Sequence 232 AA;
 SQ

Alignment Scores:
 Pred. No.: 7, 31e-87 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservative: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x ABP70043 (1-232)

QY 7 GCTCTCTGACCGAGACCATGCTTCCAGGACCACTGAGGCGGACCGTGAACGCG 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGCACCCAGGAGATG 126
 Db 22 HistyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTTCCATCTGCTTCCATCTGCTCCACTCC 186
 Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGCTCGGCACTCCCGACTACTTCAAG 246
 Db 62 CysMetTyrGlySerLysThrPheIleLysThrValSerGlyLeuProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGGCTTCACTTGGAGCGCACCCACCTACGAGACCGCGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTyrGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACACGAGACACCTCCCTGGAGCGGAGCTGCTGTGTACAAAGGTGAAGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
 QY 367 CTGGGCAACAACATCCCGCCGCGGCGGCTGATGATGACAGAACAGCCCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrGlu 141
 QY 427 CCTTCCACCGAGATCGTGTACGAGGTGGAGCGGCTGCTGGCGGCGAGTCCCTGATGGCC 486
 Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTGCCTCCCGGCTGCGCACCTGACCTGCGCACACCTACCTCCCTCAAG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTGAAAGATGCCGCGGCTTCCACTTCGAGGACCAACCGCATCGATC 606
 Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
 QY 607 CTGGAGGAGTGGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCTGGCGCGGTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGAGACCGCCCGCTCCAAAGCTGGGCGCACAAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 11
 AAG79762
 ID AAG79762 standard; protein; 232 AA.
 AC AAG79762;
 XX 01-APR-2003 (first entry)
 DT Wild type kindling fluorescent protein, AsFP595.
 XX Kindling fluorescent protein; kindling stimulus; movement; labeling;
 DE fluorescence resonance energy transfer; FRET;
 XX bioluminescence resonance energy transfer; BRET; biosensor;
 KW automated screening.
 KW Anthozoa sulcata.
 XX OS
 XX WO200296924-A1.
 XX PD
 XX 05-DEC-2002.
 XX PF
 XX 24-MAY-2002; 2002WO-US016379.
 XX PR
 XX 25-MAY-2001; 2001US-0293752P.
 XX PR
 XX 11-OCT-2001; 2001US-0329176P.
 XX PA
 XX (CLON-) CLONTECH LAB INC.
 XX PI
 XX Lukyanov SA, Chudakov D, Lukyanov K;
 XX WPI; 2003-156788/15.
 XX DR
 XX N-PSDB; ABA00799.
 XX PT
 XX Novel nucleic acid that is present in other than its natural environment
 PT and that encodes kindling fluorescent protein, is useful in labeling
 PT protocols, e.g. labeling proteins, organelles, cells and organisms.
 XX Example; Fig 1; 96pp; English.
 XX The sequences given in AAG79762-69 show wild type and mutant kindling
 CC fluorescent proteins. The proteins go from a first substantially non-

fluorescent or non-fluorescent state to a second fluorescent state upon exposure to a kindling stimulus. The kindling proteins are useful for detecting an entity such as a protein, organelle or cell in a composition such as a cell or a multicellular composition (preferably a multicellular organism), by providing the entity as an entity labeled with the kindling protein, kindling the kindling fluorescent protein label with a kindling stimulus to produce a kindling fluorescent protein label, and exciting the kindling fluorescent protein label with light and detecting any fluorescence from it to detect the entity. The method monitors the movement of the entity. The fluorescent proteins and the cDNA encoding them are useful in labeling protocols, e.g., labeling proteins, organelles, cells and organisms as biological labels or markers, in protein labeling or tagging applications. The fluorescent kindling proteins are useful as detectable labels, as labels in analyte detection assays, in fluorescence resonance energy transfer (FRET) applications, as biosensors in prokaryotic and eukaryotic cells, in applications involving the automated screening of arrays of cells expressing fluorescent reporting groups, in high through-put screening assays, as second messenger detectors, and in fluorescent activated cell sorting assays

SQ Sequence 232 AA;

Alignment Scores:
 Pred. No.: 7, 31e-87 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservatives: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: Gaps: 0

US-10-081-864-21 (1-707) x AAG79762 (1-232)

QY 7 GCTTCCCTGTCACGAGACCATGCTTCAGGACCACTGAGGCGCCGCTGACAGCGC 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValaenGly 21
 QY 67 CACTACTTCAAGTGACCGGACGAGGCGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
 Db 22 HisTyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AGATGAGTGATGAGGCGCGCCCTTCGCTTCACATCCGTGCCACCTCC 186
 Db 42 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGACGCTCCAGGCGCTTCATCAAGTACGTGCGGATCCCGACTACTTCAAG 246
 Db 62 CysMetTyroLysThrPheIleLysTyroValSerGlyIleProAspTyroPheLys 81
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGCGCACCCACCTAGGACGCGCGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyroGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACGACACCTCCCTGGAGCGGAGCTGCTGTCAGTACAGTGAGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyroLysValIle 121
 QY 367 CTGGGCAACAACTTCCCGCGACGCGCCGCTGATGCAGAACGCGCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCCACGAGATGCTGTACAGGPGAGCGGCTGCTGCGCGCGAGTCCCTGATGCC 486
 Db 142 ProAlaThrGluIleValTyroGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTCCCGCGGCTCGCACCTGACCTGCCACCTGCACACCTACCGCTCCAG 546
 Db 162 LeuLysCysProGlyArgHisLeuThrCysHisLeuHisThrThrTyroArgSerLys 181
 QY 547 AGCCCGCTCCGCTCGAGATGCGCGCTTCCACTTGGAGACCAACCGATCGATC 606
 Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201

QY 607 CTGAGGAGGTGGAGAGGCAAGTGTACAGAGTACGAGCGCGCTGGCGGTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyroLysGlnTyroGluAlaValGlyArgTyro 221
 QY 667 TCGAGCGCGCGCCCTCCCAAGCTGGGCGCACAAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 12

AAG79763

ID AAG79763 standard; protein; 232 AA.

XX AAG79763;

XX 01-APR-2003 (first entry)

XX Kindling fluorescent protein, AsP595 Al48G.

XX Kindling fluorescent protein; kindling stimulus; movement; labeling;

KW fluorescence resonance energy transfer; FRET;

KW bioluminescence resonance energy transfer; BRET; biosensor;

KW automated screening.

XX Anthozoa sulcata.

XX Key Location/Qualifiers

FT Misc-difference 148

FT /label= Al48G

XX WO200296924-A1.

XX 05-DEC-2002.

XX 24-MAY-2002; 2002WO-US016379.

XX 25-MAY-2001; 2001US-0293752P.

PR 11-OCT-2001; 2001US-0329176P.

XX (CLON-) CLONTECH LAB INC.

XX Lukyanov SA, Chudakov D, Lukyanov K;

DR WPI; 2003-156789/15.

DR N-PSDB; ABA00800.

XX Novel nucleic acid that is present in other than its natural environment

PT and that encodes kindling fluorescent protein, is useful in labeling

PT protocols, e.g. labeling proteins, organelles, cells and organisms.

XX Example; Fig 2; 96pp; English.

XX The sequences given in AAG79762-69 show wild type and mutant kindling

CC fluorescent proteins. The proteins go from a first substantially non-

CC fluorescent or non-fluorescent state to a second fluorescent state upon

CC exposure to a kindling stimulus. The kindling proteins are useful for

CC detecting an entity such as a protein, organelle or cell in a composition

CC such as a cell or a multicellular composition (preferably a multicellular

CC organism), by providing the entity as an entity labeled with the kindling

CC protein, kindling the kindling fluorescent protein label with a kindling

CC stimulus to produce a kindling fluorescent protein label, and

CC exciting the kindling fluorescent protein label with light and

CC detecting any fluorescence from it to detect the entity. The method

CC monitors the movement of the entity. The fluorescent proteins and the

CC cDNA encoding them are useful in labeling protocols, e.g., labeling

CC proteins, organelles, cells and organisms as biological labels or

CC markers, in protein labeling or tagging applications. The fluorescent

CC kindling proteins are useful as detectable labels, as labels in analyte

CC detection assays, in fluorescence resonance energy transfer (FRET)

CC applications, as biosensors in prokaryotic and eukaryotic cells, in

CC applications involving the automated screening of arrays of cells

CC expressing fluorescent reporting groups, in high through-put screening

CC assays, as second messenger detectors, and in fluorescent activated cell
 CC sorting assays
 XX Sequence 232 AA;

Alignment Scores:
 Pred. No.: 8.75e-87 Length: 232
 Score: 1216.00 Matches: 221
 Percent Similarity: 97.40% Conservative: 4
 Best Local Similarity: 95.67% Mismatches: 6
 Query Match: 93.69% Indels: 0
 DB: 6 Gaps: 0

US-10-081-864-21 (1-707) x AAG79763 (1-232)

```

QY 7 GCCTCCCTGTCAGACGACCATCCCTTCAGACACACCATCCAGGGCCCGTGAACGCG 66
DB 2 AlaserPheLeuLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTCCACCGGCAAGGGCGAGGCAACCCCTCGAGGGCAACCGAGAGATG 126
DB 22 HistyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCTTCCCTTCCACATCTGTCACCTCC 186
DB 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTAGCTGTCGGCATCCCGCACTACTTCAAG 246
DB 62 CysMetTyGlySerLysThrPheLysTyValSerGlyileProAspTyPheLys 81
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGGCGCACACACCTACGAGACGGCGCTTC 306
DB 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyGluAspGlyGlyPhe 101
QY 307 CTGACCGGCCACACGAGCACCTCCCTCGACGGGACTGCTGCTGTACAGGTGAAGATC 366
DB 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyLysVallylle 121
QY 367 CTGGGCAACAACTTCCCGCGCGAGCGCCCGTGTATGAGAACAAAGCGCGCGCTGGAG 426
DB 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCCTCCACCGAGATGCTGTACGAGGTGACGGGTGCTGCGCGCGCAGTCCCTGTATGGC 486
DB 142 ProGlyThrGluLysValTyGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGATGCCCGCGGTGCGCACCTGACCTGCGACCTGACACACCTACCGCTCCAG 546
DB 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyArgSerLys 181
QY 547 AAGCCGCTCCCGCTGAAGATCCCGGCTTCCACTTCGAGGACACCGCATCCGAGATC 606
DB 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgileGluile 201
QY 607 CTGAGGAGGTGGAGAGGGCAAGTGTACAAGCAGTACGAGCCCGCTGGCGCGCTAC 666
DB 202 MetGluGluValGluLysGlyLysCysTyLysGlnTyGluAlaAlaValGlyArgTy 221
QY 667 TGCACGCGCGCCCTCCAGCTGGGCCACAAC 699
DB 222 CysaspAlaAlaProSerLysLeuGlyHisasn 232

```

RESULT 13

AAE28834

ID AAE28834 standard; protein; 232 AA.

XX AC AAE28834;

XX DT 27-DEC-2002 (first entry)

XX DE Anemonia sulcata asPF600 (NFP-7) wild-type protein.

XX

Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 fluorescence activated cell sorting application; fluorescent timer;
 biosensor; fluorescence resonance energy transfer application; FRET;
 colouring agent; recombinant DNA application; analyte detection assay;
 sunscreener; second messenger detector; asPF600 protein; NFP-7.
 XX Anemonia sulcata.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 68
 FT /note= "Encoded by GCC"
 FT Misc-difference 143
 FT /note= "Encoded by TCC"
 XX
 XX WO200268459-A2.
 XX 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.
 XX 21-FEB-2001; 2001US-0270983P.
 XX 04-DEC-2001; 2001US-00006922.
 XX (CLON-) CLONTECH LAB INC.
 PA Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Pradkov A;
 XX WPI; 2002-691654/74.
 XX N-PSDB; AAD46279.
 XX
 XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 analyte detection assays or fluorescence activated cell sorting
 applications.
 XX Disclosure; Page 71-72; 80pp; English.

The invention relates to nucleic acid molecules encoding non-aggregating
 chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 useful in analyte detection assays, as colouring agents, as markers in
 recombinant DNA applications, as sunscreens or filters, in fluorescence
 resonance energy transfer (FRET) applications, as biosensors in
 prokaryotic and eukaryotic cells, in screening assays, as second
 messenger detectors, in fluorescence activated cell sorting applications,
 in protease cleavage assays or as fluorescent timers. The present
 sequence is Anemonia sulcata asPF600 (NFP-7) wild-type protein of the
 invention

SQ Sequence 232 AA;

Alignment Scores:

Pred. No.: 1-25e-86 Length: 232
 Score: 1216.00 Matches: 220
 Percent Similarity: 97.84% Conservative: 6
 Best Local Similarity: 95.24% Mismatches: 5
 Query Match: 93.54% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x AAE28834 (1-232)

```

QY 7 GCCTCCCTGTCAGACGACCATCCCTTCAGACACACCATCCAGGGCCCGTGAACGCG 66
DB 2 AlaserPheLeuLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTCCACCGGCAAGGGCGAGGCAACCCCTCGAGGGCAACCGAGAGATG 126
DB 22 HistyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCTTCCCTTCCACATCTGTCACCTCC 186
DB 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTAGCTGTCGGCATCCCGCACTACTTCAAG 246

```

Db 62 CysMetTyGlySerLysThrPheIleLysTyValSerGlyIleProAspTyPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTACCTGGGAGCGCCACCACTACGAGCGCGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACAGGACACTCCCTGCAGCGGACTGCTCGGTGTACAAGGTGAAGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyLysValLysIle 121
 QY 367 CTGGCAACAACCTTCCCGCGAGCGCCCGTGTATGAGAACAGCGCGCGTGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 142 ProAlaThrGluIleValTyGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGTGCCTCCCGCGGTCGCACTGACCTGCGCAGCTGCACACCACTACCGCTCAAG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTGAAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
 Db 182 LysProAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
 QY 607 CTGAGGAGTGGAGAGGCAAGTGTACAGCAGTACAGCGCGCGCGTGGCCCTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyLysGlnTyGluAlaValGlyArgTy 221
 QY 667 TGCAGCCCGCCCTCCAGCTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 14

AAG79764
 ID AAG79764 standard; protein; 232 AA.

AC AAG79764;

DT 01-APR-2003 (first entry)

XX Kindling fluorescent protein, AsFP595 F90L, A148G, H203Y.

XX Kindling fluorescent protein; kindling stimulus; movement; labeling;

KW fluorescence resonance energy transfer; FRET;

KW bioluminescence resonance energy transfer; BRET; biosensor;

XX automated screening.

XX Anthozoa sulcata.

XX Key Location/Qualifiers

FT Misc-difference 90

FT FT /label= F90L

FT FT /label= A148G

FT FT /label= H203Y

XX WO200296924-A1.

XX 05-DEC-2002.

XX 24-MAY-2002; 2002WO-US016379.

XX 25-MAY-2001; 2001US-0293752P.

XX 11-OCT-2001; 2001US-0329176P.

XX (CLON-) CLONTECH LAB INC.

XX Lukyanov SA, Chudakov D, Lukyanov K;

XX

DR WPI; 2003-156788/15.
 DR N-PSDB; ABA00801.
 XX Novel nucleic acid that is present in other than its natural environment
 PT and that encodes kindling fluorescent protein, is useful in labeling
 PT protocols, e.g. labeling proteins, organelles, cells and organisms.
 XX Example; Fig 3; 96pp; English.
 PS The sequences given in AAG79762-69 show wild type and mutant kindling
 CC fluorescent proteins. The proteins go from a first substantially non-
 CC fluorescent or non-fluorescent state to a second fluorescent state upon
 CC exposure to a kindling stimulus. The kindling proteins are useful for
 CC detecting an entity such as a protein, organelle or cell in a composition
 CC such as a cell or a multicellular composition (preferably a multicellular
 CC organism), by providing the entity as an entity labeled with the kindling
 CC protein, kindling the kindling fluorescent protein label with a kindling
 CC stimulus to produce a kindling fluorescent protein label with a kindling
 CC exciting the kindling fluorescent protein label with light and
 CC detecting any fluorescence from it to detect the entity. The method
 CC monitors the movement of the entity. The fluorescent proteins and the
 CC cDNA encoding them are useful in labeling protocols, e.g., labeling
 CC proteins, organelles, cells and organisms, as biological labels or
 CC markers, in protein labeling or tagging applications. The fluorescent
 CC kindling proteins are useful as detectable labels, as labels in analyte
 CC detection assays, in fluorescence resonance energy transfer (FRET)
 CC applications, in bioluminescence resonance energy transfer (BRET)
 CC applications, as biosensors in prokaryotic and eukaryotic cells, in
 CC applications involving the automated screening of arrays of cells
 CC expressing fluorescent reporting groups, in high through-put screening
 CC assays, as second messenger detectors, and in fluorescent activated cell
 CC sorting assays
 XX
 SQ Sequence 232 AA;

Alignment Scores:

Pred. No.: 7,59e-86 Length: 232
 Score: 1206.00 Matches: 219
 Percent Similarity: 96.97% Conservative: 5
 Best Local Similarity: 94.81% Mismatches: 7
 Query Match: 92.77% Indels: 0
 DB: 6 Gaps: 0

US-10-081-864-21 (1-707) x AAG79764 (1-232)

QY 7 GCCTCCCTGTGACCGAGACCATGCTTCAGGACCAACATCGAGGCAACCGTGAACGGC 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGCAAGGGGCAACCCCTCGAGGCGCACCGAGGAGATG 126
 Db 22 HistyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGTGTATGAGGCGCGCCCTTCCTTCCTTCACATCTCTGTCCACCTCC 186
 Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCCAAGGCTTCATCAAGTGTGTCGGGATCCCGCACTACTTCAAG 246
 Db 62 CysMetTyGlySerLysThrPheIleLysTyValSerGlyIleProAspTyPheLys 81
 QY 247 CACTCCCTCCCGAGGGCTTCACCTGGGAGCGCCACCCACCTACGAGACCGCGCTTC 306
 Db 82 GlnSerPheProGluGlyLeuThrTrpGluArgThrThrTyGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACAGGACCACTCCCTGGAGCGGAGTGTCTGTGTACAGGTGAAGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyLysValLysIle 121
 QY 367 CTGGGCAACAACCTTCCCGCGAGCGCCCGTGTATGAGCAACAGCGCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141

QY 427 CCCTCCACCGAGATCGTGTAGAGGTGGACGGGTGCTGGCGGCCAGTCCCTGTATGGCC 486
 DB 142 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTCCCGCGGGTCCGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 546
 DB 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTCGAAGATCCCGCTTCACCTTCGAGGACCAACCGCATCGAGATC 606
 DB 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspTyrArgIleGluIle 201
 QY 607 CTGGAGGAGGTGGAGAGGCAAGTGTACAGCAGTACAGCAGTACAGCAGTACAGCAGTAC 666
 DB 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGACGCGCGCCCTCCCAAGCTGGGCCACAAC 699
 DB 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 15
 AAG79467
 ID AAG79467 standard; protein; 226 AA.
 XX
 AC AAG79467;
 DT 29-NOV-2002 (first entry)
 XX
 DE C. gigantea mutant chromoprotein.
 XX
 KW Chromoprotein; CgCP; C. gigantea; fluorescent; emission maximum;
 KW labeling; colouring agent; food; pharmaceutical; cosmetic; biosensor;
 KW fluorescence resonance energy transfer; detection assay; FRET;
 KW selectable marker.
 XX
 OS Condylactis gigantea.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 142
 FT /label= C148S
 FT /note= "Numbering according to GFP"

WO200259309-A2.
 01-AUG-2002.
 11-DEC-2001; 2001WO-US047995.
 13-DEC-2000; 2000US-0255533P.
 (CLON-) CLONTECH LAB INC.
 PI Lukyanov SA, Lukyanov KA, Fradkov AF;
 XX WPI; 2002-666902/71.
 DR N-PSDB; ABA00245.
 PT New nucleic acids encoding chromoproteins or fluorescent proteins, useful
 PT as labeling tools for marking a protein, cell or organism, which may be
 PT used in biochemistry, molecular biology or medical diagnostic
 PT applications.
 XX
 PS Disclosure; Fig 2; 56pp; English.
 XX

The sequences given in AAG79466-67 represent wild type and mutant
 CC chromoprotein (CgCP) from C. gigantea. The mutant protein is fluorescent
 CC having an emission maximum ranging from about 480-680 nm. The wild type
 CC and mutant proteins are useful as labeling tools for marking a protein,
 CC cell or organism, which may be used in biochemistry, molecular biology or
 CC medical diagnostic applications. The chromoproteins may be used as
 CC colouring agents capable of imparting colour or pigment to a particular
 CC composition of matter, e.g. food compositions, pharmaceuticals or
 CC cosmetics. They may also be used as labels in analyte detection assays,
 CC

CC or as selectable markers in recombinant DNA applications. The fluorescent
 CC proteins may be used in fluorescence resonance energy transfer (FRET)
 CC applications, e.g. detection of protein-protein interactions. They are
 CC also used as biosensors in prokaryotic and eukaryotic cells, e.g. as pH
 CC indicator or phosphorylation indicator. The proteins are useful as in
 CC vivo marker in animals, in assays to determine the phospholipid
 CC composition in biological membranes, or in protease cleavage assays
 XX
 SQ Sequence 226 AA;
 Alignment Scores:
 Pred. No.: 4,646-53 Length: 226
 Score: 786.50 Matches: 149
 Percent Similarity: 78.85% Conservative: 30
 Best local Similarity: 65.64% Mismatches: 45
 Query Match: 60.50% Indels: 3
 DB: 5 Gaps: 3
 US-10-081-864-21 (1-707) x AAG79467 (1-226)

QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCACTCGAGGCGCCGTTGACCGC 66
 DB 1 AlaGlyLeuLeuLysGluSerMetArgIleLysIleTyrMetGluGlyThrValAsnGly 20
 QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGCGCCACCGAGAGATG 126
 DB 21 TyrHisPheLysCysGluGlyAspGlyAsnPropheGluGlyThrGlnAsnMet 40
 QY 127 AAGATCGAGTGTATCGAGGGCGGCCCTCGCTTCGCTCCACATCGTCTCCACTCC 186
 DB 41 ArgIleArgValThrGluGlyAlaProLeuPropheAlaPheAspIleLeuSerProCys 60
 QY 187 TGCATGTACGGCTCCCAAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGACCTACTTCAAG 246
 DB 61 CysAlaTyrGlySerLysThrPheLysHisThrSerGlyIleProAspTyrPheLys 80
 QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGCACCACTACGAGGCGCGGCTTC 306
 DB 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrIleTyrGluAspGlyGlyVal 100
 QY 307 CTGACCGCGCCACCGAGCACCTCCCTGGAGCGGACTGCTGTGTACAGGTGAAGATC 366
 DB 101 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnCysLeuIleTyrLysVal 120
 QY 367 CTGGGCAACACTTCCCGCGGCGCGCTTCATGATGACAGACAGCGCGCGCTCGGAG 426
 DB 121 LeuGlyThrAsnPheProAlaAspGlyProValMetLysLysIleSerGlyGlyTrpGlu 140
 QY 427 CCCTCCACCGAGATCGGTGTACGAGGTGGACGGCGTGTGCGCGCGCAGTCCCTGTATGGCC 486
 DB 141 ProSerThrGluIleValTyrGlnAspAsnGlyValLeuArgGlyArgAsnValMetAla 160
 QY 487 CTGGAGTCCCGCGCGCTCCGACCTGACCTGCACCTGCACACCACTACCGCTCCAG 546
 DB 161 LeuLysValSerGlyArgProProLeuIleCysHisLeuHisSerThrTyrArgSerLys 180
 QY 547 AAGCCCGCTCCCGCTGGAAGATGCCCGCTTCACCTTCGAGGACCAACCGCATCGAGATC 606
 DB 181 Lys---AlaCysAlaLeuThrMetProGlyPheHisPheAlaAspLeuArgIleGln--- 198
 QY 607 CTGGAGGAGGTGGAGAGGCAAGTGTCTACAGCAGTACAGGCGCGCGCTCGGCGCTAC 666
 DB 199 MetProLysLysLysLysAspGluTyrPheGluLeuTyrGluAlaSerValAlaArgTyr 218
 QY 667 TGGACGCGCGCCCTCCCAAG 687
 DB 219 SerAsp---ValProGluLys 224

Search completed: July 29, 2004, 14:32:30
 Job time : 54.2812 secs

Blank Sheet USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:29:01 ; Search time 14.5484 Seconds
(without alignments)
5017.682 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300

Sequence: 1 ggatccgcctcctcgtgac.....ctgggcacacactgaagctt 707

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_P/US10081864/runat_23072004_150507_25519/app_query.fasta_1.1742
-DB=Issued Patents AA -QFMT=faстан -SUFFIX=rai -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=Eits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864@cgn1.1.45 @runat_23072004_150507_25519 -RCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/6C.COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453.5	34.9	238	3	US-09-277-716-32
2	453.5	34.9	238	4	US-09-609-161B-32
3	445.5	34.3	238	3	US-09-277-716-16
4	445.5	34.3	238	4	US-09-609-161B-16
5	445.5	34.3	238	4	US-09-626-581B-65
6	445.5	34.3	238	4	US-09-415-765B-65
7	445.5	34.3	238	4	US-09-626-580C-65
8	433.5	33.3	238	4	US-09-839-650-3
9	239	17.8	377	4	US-09-252-991A-31793
10	232.5	17.9	247	4	US-09-252-991A-26899
11	228.5	17.6	238	4	US-09-472-065A-5
12	228.5	17.6	997	4	US-09-417-197-121

13	227	17.5	302	4	US-09-252-991A-19529	Sequence 19529, A
14	225.5	17.3	631	4	US-09-417-197-39	Sequence 39, Appl
15	225	17.3	322	4	US-09-252-991A-31608	Sequence 31608, A
16	224.5	17.3	238	4	US-09-023-946B-20	Sequence 20, Appl
17	224.5	17.3	238	4	US-09-023-946B-30	Sequence 30, Appl
18	224.5	17.3	238	4	US-09-023-946B-31	Sequence 31, Appl
19	224.5	17.3	238	4	US-09-023-946B-32	Sequence 32, Appl
20	224.5	17.3	238	4	US-09-023-946B-33	Sequence 33, Appl
21	224.5	17.3	238	4	US-09-023-946B-37	Sequence 37, Appl
22	224.5	17.3	1070	4	US-09-091-042A-2	Sequence 2, Appl
23	224	17.2	843	4	US-09-417-197-117	Sequence 117, App
24	223.5	17.2	238	3	US-08-643-704A-49	Sequence 49, Appl
25	223.5	17.2	242	4	US-09-023-946B-1	Sequence 1, Appl
26	222.5	17.1	218	4	US-09-603-448-5	Sequence 5, Appl
27	222.5	17.1	238	1	US-08-452-295-1	Sequence 1, Appl
28	222.5	17.1	238	4	US-09-213-343-4	Sequence 4, Appl
29	222	16.6	197	4	US-09-252-991A-26537	Sequence 26537, A
30	222	17.1	1171	4	US-09-417-197-131	Sequence 131, App
31	220.5	17.0	239	3	US-09-094-359-10	Sequence 10, Appl
32	220.5	17.0	239	3	US-09-172-083-5	Sequence 5, Appl
33	220.5	17.0	239	4	US-09-316-919-6	Sequence 6, Appl
34	220.5	17.0	239	4	US-09-602-641-5	Sequence 5, Appl
35	220.5	17.0	239	4	US-09-704-463-10	Sequence 10, Appl
36	219.5	16.9	238	4	US-09-023-946B-24	Sequence 24, Appl
37	219	16.8	941	4	US-09-513-783A-172	Sequence 172, App
38	218.5	16.8	238	3	US-08-893-327-16	Sequence 16, Appl
39	218.5	16.8	238	4	US-09-472-065A-6	Sequence 6, Appl
40	218.5	16.8	239	3	US-09-094-359-4	Sequence 4, Appl
41	218.5	16.8	239	3	US-09-172-063-3	Sequence 3, Appl
42	218.5	16.8	239	4	US-09-513-783A-46	Sequence 46, Appl
43	218.5	16.8	239	4	US-09-316-919-4	Sequence 4, Appl
44	218.5	16.8	239	4	US-09-602-641-3	Sequence 3, Appl
45	218.5	16.8	239	4	US-09-704-463-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-277-716-32

; Sequence 32, Application US/09277716A

; Patent No. 6232107

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLINE, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

; CURRENT APPLICATION NUMBER: US/09/277,716A

; CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: 60/102,939

; EARLIER FILING DATE: 1998-10-01

; EARLIER APPLICATION NUMBER: 60/089,367

; EARLIER FILING DATE: 1998-06-15

; EARLIER APPLICATION NUMBER: 60/079,624

; EARLIER FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 32

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Ptilosarcus gurneyi

; FEATURE:

; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)

US-09-277-716-32

Alignment Scores:			
Pred. No.:	6,548-29	Length:	238
Score:	453.50	Matches:	94
Percent Similarity:	58.33%	Conservative:	32
Best Local Similarity:	43.52%	Mismatches:	87
Query Match:	34.88%	Indels:	3
DB:	3	Gaps:	2

US-10-081-864-21 (1-707) x US-09-277-716-32 (1-238)

```

;
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-609-161B-32

Alignment Scores:
Pred. No.: 6,548-29 Length: 238
Score: 453.50 Matches: 94
Percent Similarity: 58.33% Conservative: 32
Best Local Similarity: 43.53% Mismatches: 87
Query Match: 34.88% Indels: 3
DB: 4 Gaps: 2

US-10-081-864-21 (1-707) x US-09-609-161B-32 (1-238)

QY 16 CTGACCGAGACCATGCTCCCTCAGGACCAACCCCTCAGGAGCCCGTGAACGGCCACTACTTC 75
Db 11 LeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsnHisValPhe 30

QY 76 AAGTGCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 135
Db 31 SerMetGluGlyPheGlyGlyValLeuPheGlyGlyValLeuPheGlyGlyValLeuPhe 50

QY 136 GTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIleAlaPheGlnTyr 70

QY 196 GGTCTCAAGGCGTTCATCAAGTACGTGTCGGGATCCCGGACTCTTCAAGCAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrLysTyrProAspIleAlaAspTyrPheValGlnSerPhe 90

QY 256 CCGGAGGCGTTCACCTGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 315
Db 91 ProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAlaIleValAspIle 110

QY 316 CACGAGGACACTCCCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 375
Db 111 ArgSerAspIleSerLeuGluAspLysPheHisTyrLysValGluTyrArgGlyAsn 130

QY 376 AACTTCCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
Db 131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150

QY 436 GAGATCGTGTACGAGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
Db 151 GluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeuValTyrLysLeu 170

QY 496 CCGGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
Db 171 GluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSerLys-----Gly 188

QY 556 TCCGCGCTGAAGATGCCCGGCTCCACTTCGAGGACCAACCGCATCCGAGGAGGAGGAGGAG 615
Db 189 GlyValLysGluPheProGluTyrHisPheIleHisArgLeuGlu-----LysThrTyr 207

QY 616 GTGAGAGGCGCAAGTGTCTACAGCAGTACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 663
Db 208 ValGluGluGlySerPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 2
US-09-609-161B-32
; Sequence 32, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi

```

```

;
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-609-161B-32

Alignment Scores:
Pred. No.: 6,548-29 Length: 238
Score: 453.50 Matches: 94
Percent Similarity: 58.33% Conservative: 32
Best Local Similarity: 43.53% Mismatches: 87
Query Match: 34.88% Indels: 3
DB: 4 Gaps: 2

US-10-081-864-21 (1-707) x US-09-609-161B-32 (1-238)

QY 16 CTGACCGAGACCATGCTCCCTCAGGACCAACCCCTCAGGAGCCCGTGAACGGCCACTACTTC 75
Db 11 LeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsnHisValPhe 30

QY 76 AAGTGCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 135
Db 31 SerMetGluGlyPheGlyGlyValLeuPheGlyGlyValLeuPheGlyGlyValLeuPhe 50

QY 136 GTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIleAlaPheGlnTyr 70

QY 196 GGTCTCAAGGCGTTCATCAAGTACGTGTCGGGATCCCGGACTCTTCAAGCAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrLysTyrProAspIleAlaAspTyrPheValGlnSerPhe 90

QY 256 CCGGAGGCGTTCACCTGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 315
Db 91 ProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAlaIleValAspIle 110

QY 316 CACGAGGACACTCCCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 375
Db 111 ArgSerAspIleSerLeuGluAspLysPheHisTyrLysValGluTyrArgGlyAsn 130

QY 376 AACTTCCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
Db 131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150

QY 436 GAGATCGTGTACGAGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
Db 151 GluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeuValTyrLysLeu 170

QY 496 CCGGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
Db 171 GluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSerLys-----Gly 188

QY 556 TCCGCGCTGAAGATGCCCGGCTCCACTTCGAGGACCAACCGCATCCGAGGAGGAGGAGGAG 615
Db 189 GlyValLysGluPheProGluTyrHisPheIleHisArgLeuGlu-----LysThrTyr 207

QY 616 GTGAGAGGCGCAAGTGTCTACAGCAGTACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 663
Db 208 ValGluGluGlySerPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 3
US-09-277-716-16
; Sequence 16, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi

```

EARLIER APPLICATION NUMBER: 60/079,624

EARLIER FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 16

LENGTH: 238

TYPE: PRT

ORGANISM: Renilla mulleri

FEATURE:

OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Alignment Scores:
Pred. No.: 238
Score: 445.50
Percent Similarity: 59.26%
Best Local Similarity: 42.58%
Query Match: 34.27%
DB: 3
Gaps: 2

US-10-081-864-21 (1-707) x US-09-277-716-16 (1-238)

```
QY 16 CTGACCGAGACCATCCCTTCAGGACCCACCATCGAGGCGACCGTCAAGCGCCACTACTTC 75
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
DB 31 ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGGCGGCGGCGGCTCGCTCCACATCCCTCGACCTCCCTCGATGATAC 195
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGTCTCAAGCGCTTCATCAAGTACGTCTCGGCGATCCCGACTACTTCAAGCGTCCCTC 255
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
QY 256 CCGAGGGTTCACCTGGAGCGGCGGCTCGGCGGCGGCTCGGCGGCGGCTCGGCGGCGG 315
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
QY 316 CACGACGACCTCCCTGGAGCGGCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
DB 111 ArgSerAspIleAsnLeuLeuGluValGluValGluValGluValGluValGluVal 130
QY 376 AACTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
DB 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTGTACGAGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValGlyGluValGlyGluVal 170
QY 496 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
DB 171 AsnSerGlyLysTyrSerCysHisMetLysThrLysMetLysSerLys-----Gly 188
QY 556 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 615
DB 189 ValValLysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663
DB 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
```

RESULT 4

US-09-609-161B-16

Sequence 16, Application US/09609161B

Patent No. 6436682

GENERAL INFORMATION:

APPLICANT: Bryan, Bruce

APPLICANT: Szent-Gyorgyi, Christopher

APPLICANT: PROLUME, LTD.

TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC

TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG

FILE REFERENCE: 24729-121B

CURRENT APPLICATION NUMBER: US/09/609,161B

CURRENT FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: 09/277,716

PRIOR FILING DATE: 1999-03-26

PRIOR APPLICATION NUMBER: 60/102,939

PRIOR FILING DATE: 1998-10-01

PRIOR APPLICATION NUMBER: 60/089,367

PRIOR FILING DATE: 1998-06-15

PRIOR APPLICATION NUMBER: 60/079,624

PRIOR FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 16

LENGTH: 238

TYPE: PRT

ORGANISM: Renilla mulleri

FEATURE:

OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-609-161B-16

Alignment Scores:
Pred. No.: 238
Score: 445.50
Percent Similarity: 59.26%
Best Local Similarity: 42.58%
Query Match: 34.27%
DB: 3
Gaps: 2

US-10-081-864-21 (1-707) x US-09-609-161B-16 (1-238)

```
QY 16 CTGACCGAGACCATCCCTTCAGGACCCACCATCGAGGCGACCGTCAAGCGCCACTACTTC 75
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
DB 31 ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGGCGGCGGCGGCTCGCTCCACATCCCTCGACCTCCCTCGATGATAC 195
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGTCTCAAGCGCTTCATCAAGTACGTCTCGGCGATCCCGACTACTTCAAGCGTCCCTC 255
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
QY 256 CCGAGGGTTCACCTGGAGCGGCGGCTCGGCGGCGGCTCGGCGGCGGCTCGGCGGCGG 315
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
QY 316 CACGACGACCTCCCTGGAGCGGCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
DB 111 ArgSerAspIleAsnLeuLeuGluValGluValGluValGluValGluValGluVal 130
QY 376 AACTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
DB 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTGTACGAGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValGlyGluValGlyGluVal 170
QY 496 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
DB 171 AsnSerGlyLysTyrSerCysHisMetLysThrLysMetLysSerLys-----Gly 188
QY 556 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 615
DB 189 ValValLysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663
DB 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
```

Db 189 ValVallysGluPheProSerTyHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGAGGCAAGTGTACAAAGCAGTACAGAGCCCGCTGGCGCC 663
Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
RESULT 5
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-581D-65
Alignment Scores:
Pred. No.: 2,98-28 Length: 238
Score: 445.50 Matches: 92
Percent Similarity: 59.26% Conservative: 36
Best Local Similarity: 42.59% Mismatches: 85
Query Match: 34.27% Indels: 3
Gaps: 2

US-10-081-864-21 (1-707) x US-09-626-581D-65 (1-238)
QY 16 CTGACGAGACCATGCGCTTCAGGAGCACCACCATCGAGGCGCCGCTGAACGGCCACTCTTC 75
Db 11 LeuGlnGluValMetSerTyHisValAsnLeuGluGlyLeuValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATCGAG 135
Db 31 ThrMetGluGlyCysGlyGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGCGGCGCCCTCGCTTCGCTTCACATCTGTGTCACCTCTCGCATGTAC 195
Db 51 ValThrlysglyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGTCTCAAGGCGCTTCATCAAGTACGTGTCGGGATCCCGGATCTTCAAGCAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrlystyrProAsnAspIleSerAspTyHisPheIleGlnSerPhe 90
QY 256 CCGAGGCGCTTCACCTCGGAGCGCACCCACCATCGAGGCGGCGCTTCCTGACCGCC 315
Db 91 ProAlaGlyPheMetTyHisValArgThrLeuArgTyHisValGlyLeuValGluIle 110
QY 316 CACCAGGACACCTCCCTGGAGCGGCGCTGCTGTACAAAGTGAAGATCTCTGGGCAAC 375
Db 111 ArgSerAspIleAsnLeuIleGluAspPheValTyHisPheValGlyLeuValGlySer 130
QY 376 AACTTCCCGCGGCGGCGCTGTGTCAGAGCAAGCGCGCGCTGGGAGCGCTCCACCC 435
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTACGAGTGAAGCGGCTGCTGGCGGCGGCGCTGCTGATGCGGCGCTGGAGTGC 495
Db 151 GluAlaMetTyHisMetAsnAsnGlyValLeuValGlyLeuValTyHisLeu 170
QY 496 CCGCGGCGTCCGACCTGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555

Db 171 AsnSerGlyLysTyHisPheIleGlnHisArgLeuGlu---LysThrTyr 188
QY 556 TCCGCGCTGAGATGCGCGGCTTCACCTTCGAGGAGCACCAGCAGATCTCGAGAGAG 615
Db 189 ValVallysGluPheProSerTyHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGAGGCAAGTGTCTCAAGCAGTACAGGCGCGCTGGCGCCG 663
Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
RESULT 6
US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-415-765B-65
Alignment Scores:
Pred. No.: 2,98-28 Length: 238
Score: 445.50 Matches: 92
Percent Similarity: 59.26% Conservative: 36
Best Local Similarity: 42.59% Mismatches: 85
Query Match: 34.27% Indels: 3
Gaps: 2

US-10-081-864-21 (1-707) x US-09-415-765B-65 (1-238)
QY 16 CTGACGAGACCATGCGCTTCAGGAGCACCACCATCGAGGCGCCGCTGAACGGCCACTCTTC 75
Db 11 LeuGlnGluValMetSerTyHisValAsnLeuGluGlyLeuValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATCGAG 135
Db 31 ThrMetGluGlyCysGlyGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGCGGCGCCCTCGCTTCGCTTCACATCTGTGTCACCTCTCGCATGTAC 195
Db 51 ValThrlysglyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGTCTCAAGGCGCTTCATCAAGTACGTGTCGGGATCCCGGATCTTCAAGCAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrlystyrProAsnAspIleSerAspTyHisPheIleGlnSerPhe 90
QY 256 CCGAGGCGCTTCACCTCGGAGCGCACCCACCATCGAGGCGGCGCTTCCTGACCGCC 315
Db 91 ProAlaGlyPheMetTyHisValArgThrLeuArgTyHisValGlyLeuValGluIle 110
QY 316 CACCAGGACACCTCCCTGGAGCGGCGCTGCTGTACAAAGTGAAGATCTCTGGGCAAC 375
Db 111 ArgSerAspIleAsnLeuIleGluAspPheValTyHisPheValGlyLeuValGlySer 130
QY 376 AACTTCCCGCGGCGGCGCTGTGTCAGAGCAAGCGCGCGCTGGGAGCGCTCCACCC 435
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTACGAGTGAAGCGGCTGCTGGCGGCGGCGCTGCTGATGCGGCGCTGGAGTGC 495
Db 151 GluAlaMetTyHisMetAsnAsnGlyValLeuValGlyLeuValTyHisLeu 170

QY 436 GAGATCGTGTACGAGGTGGACGGCGTGTCTGCGCGGCACGATCCCTGATGCGCCTTGGAGTGC 495
 Db 151 GluAlaMetTyMetAsnAsnGlyValLeuValGlyGluValIleLeuValTyLysLeu 170
 QY 496 CCGCGGGTGGCCACCTGACCTGGCCACCTGCACACCTACCGCTCCAAGAAGCCCGCC 555
 Db 171 AsnSerGlyIlySerTyTySerCysHisMetIlyThrLeuMetIlySerLys-----Gly 198
 QY 556 TCGCCCTCGAAGATCCCGGTCTTCACCTCGAGGACCAACCGCATCGAGATCTCTGGAGGAG 615
 Db 189 ValValIlySGlyPheProSerTyHisPheIleGlnHisargLeuGlu---LysThrTy 270
 QY 616 GTGCGAAGCGCAAGTGCTACAAAGCAGTACGAGCCCGCTGGGCGCC 663
 Db 208 ValGluAspGlyClyPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 8
 US-09-839-650-3
 ; Sequence 3, Application US/09839650
 ; Patent No. 6645761
 ; GENERAL INFORMATION:
 ; APPLICANT: Stratagene
 ; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mull
 ; Patent No. 6645761
 ; TITLE OF INVENTION: Fluorescent Protein
 ; FILE REFERENCE: 25436/1755
 ; CURRENT APPLICATION NUMBER: US/09/839,650
 ; CURRENT FILING DATE: 2001-04-19
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 238
 ; TYPE: PRP
 ; ORGANISM: Renilla muellei

Alignment Scores:					
Pred. No.:	2,71e-27	Length:	238		
Score:	433.50	Matches:	91		
Percent Similarity:	58.33%	Conservative:	35		
Best Local Similarity:	42.13%	Mismatches:	87		
Query Match:	33.35%	Indels:	3		
DB:	4	Gaps:	2		
US-10-081-864-21 (1-707) x US-09-839-650-3 (1-238)					
Qy	16	CTGACCGAGACCATGCCCTTCAGGACACCATCGAGGGCACCGTGAACGGCCACTACTTC	75		
Db	11	LeuGlnGluValMetSerTyrIlysValAsnLeuGluGlyIleValAsnHisValPhe	30		
Qy	76	AACTGCACCGCAAGGGGGGACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG	135		
Db	31	ThrMetGluGlyCysArgLysArgAsnIleLeuPheGlyAsnGlnLeuValHisIleArg	50		
Qy	136	GTGATCGAGGGCGGCGCCCTCGCTTCGCCCTTCACATCTCTGTCCACTCTCTCATGTAC	195		
Db	51	ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerProAlaPheGlnIyr	70		
Qy	196	GGCTCCAAAGCGCTTCATCAAGTACGTGTCGGCATCCCGACTACTTTCAGAGCATGCCCTC	255		
Db	71	GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe	90		
Qy	256	CCCGAGGGCTTCACCTGGAGCGCACACACCTACGAGGACGGCGGCTCTCTGACCGCC	315		
Db	91	ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle	110		
Qy	316	CACCAAGGACACTCCCTGGACGGCGACCTCGTGTGTACAGGTAAGATCTCTGGGCAAC	375		
Db	111	ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer	130		
Qy	376	AACTTCCCGCGAGGCGCCCGTGTGATCGAGAACAAAGCGCGGCGCTGGAGCGCCTCTCAC	435		


```

Db 131 AsnPheProAspAspGlyProValMetGlnLysThrLeuGlyLeuGluProSerPhe 150
QY 436 GAGATCGGTACGAGGTGGCGGTGTCGGCGCGCAGTCCCTGATGGCCCTGGAGTGC 495
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValLeuValTyrLysLeu 170
QY 496 CCCGGCGTCCACCTGACCTGACCTGCACCTGCACACCACTACCGCTCCAGAGCCGCC 555
Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLys-----Gly 188
QY 556 TCCGCCCTGAGATGCCCGGTTCCTACTTCGAGGACCGCAGCATCGAGATCCTGGAGGAG 615
Db 189 ValValLysGluPheProSerTyrHisPheLeuGlnHisArgLeuGlu-----LysThrTyr 207
QY 616 GTGGAAGAGGCAAGTGTACAGCAGTACGAGCGCCCGTGGCGCGC 663
Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaLeuAlaGln 223

RESULT 9
US-09-252-991A-31793
; Sequence 31793, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31793
; LENGTH: 377
; TYPE: PPT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (59)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-31793

Alignment Scores:
Pred. No.: 1.64e-11 Length: 377
Score: 239.00 Matches: 96
Percent Similarity: 34.86% Conservative: 18
Best Local Similarity: 29.36% Mismatches: 89
Query Match: 17.82% Indels: 124
DB: 4 Gaps: 20

US-10-081-864-21 (1-707) x US-09-252-991A-31793 (1-377)
QY 700 AGTTGTGCCAGCTGTGAGCGGCGG----- 674
Db 37 AlaCysAlaProAlaGlyAlaGlyArgGlyTyrLeuProAlaArgPheArgArg 56
QY 673 -----CGTCCGACGT----- 665
Db 57 AlaPro***AlaGlyArgArgArgGlyArgProAlaLeuArgSerAlaValAlaArg 76
QY 664 ---AGCGGCCACCGCGCCCT-----CGTACTGCTTGT 635
Db 77 ArgAlaGlyAlaArgArgProGlySerHisGlyArgArgSerAlaGlyArgValAlaAla 96
QY 634 AGCACTTCCCTTCCACCTCCCTCCAGATCTCGATCGGTGGT-----CCTCGA 584
Db 97 GlyArgAlaProSer---AlaAlaProGlySerArgLeuGlyAlaValArgProAla 115
QY 583 AGTGAAGCCGGCATCTTCAGGCGGAGCGGGGCTTCTTGGAGCGGTAGTGTGTGCA 524
Db 116 AlaAlaProArgCysArgAsnAlaProArgAlaGly-----ProCysGlu 131

```

```

QY 523 GGTGCGAG-----TCAGTGGCGACCGCCGCGGCGACTCCAGGG 485
Db 132 ArgGlyArgAlaThrProProArgAspAlaAlaGlyTyrArgArgSerSerAlaPro 151
QY 484 CCATCAGGAGTGGCGCGCAGCAGCGCTCCCTCCACCATCTCGGTGGAGGGCT 425
Db 152 AlaAlaAlaAlaGlyArgSerArgAlaProProValArgAspThrTyrProAla 171
QY 424 CCC---AGCGGCGCGCTTGTCTCATCACGGGCGCTCGCGGGGAAGT----- 377
Db 172 ProValArgGlyArgSerArgAlaArgArgAlaArgAlaThrGlyArgPro 191
QY 376 -----TGTTGCCAGGATCTTCACCTTGT 353
Db 192 HisTyrLeuProValArgAsnAlaAlaArgAlaLeuArgCysCysProAlaThrAlaProCys 211
QY 352 ACACCGAGCGAGTCGC-----CGTCAGGAGGTGTCTCTGGTGGCGGTCA 308
Db 212 ValProGlyHisArgAspAlaArgArgGlyAlaProGlyArg---ProGlySerArgAla 230
QY 307 GGAAGC-----CGCGTCTCT 293
Db 231 GlyArgTyrArgLeuProGlyArgAlaCysArgArgValProAlaAlaLeuArgArgPro 250
QY 292 CGTAGTGGTGGTGGCTCCCGAGGTGAAGCCCTCGGCGAGGAGCTGTGAAGT----- 239
Db 251 ProThrAlaTyr---ArgProArg-----ArgGlyGlyTyrCysTyrProGlyAspGly 266
QY 238 -----AGTCGGGATGCCGACACGCTACTGTGAAGGCT 203
Db 267 ArgLeuArgArgIleArgGlyArgGlySerGlyCysAlaAla----- 280
QY 202 TCGAGCCGTACATCGCAGGAGTGGACAGATGTGGAAGCGAGGCGAGCGCGC--- 146
Db 281 ---AlaArgSerCysArgTyrTyr-----CysGlyArgProArgProGlySerAlaAla 297
QY 145 -----CCTCGATCACCTCGATCTTCATCTCTCGGTGCTCGAGGGGGTTCCT 95
Db 298 GlnSerTyrProValSerSerGlnAlaSerGlyGluValSerProCysGlySerCysArg 317
QY 94 CGC-----CCTTGGCGGTGCATTTGAAGTAGTGCCTTCACGTCGCTCGATGG 44
Db 318 ArgSerArgAlaAlaCysGlnSerProLysArgSerLeuSerSerAlaSerProGly--- 336
QY 43 TGTCTCTGAAGCGCATGCT 23
Db 337 -----GlyAlaTyrSer 340

RESULT 10
US-09-252-991A-26899
; Sequence 26899, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26899
; LENGTH: 247
; TYPE: PPT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26899

Alignment Scores:
Pred. No.: 4.92e-11 Length: 247

```

Score: 232.50 Matches: 88
Percent Similarity: 40.39% Conservative: 15
Best Local Similarity: 34.51% Mismatches: 79
Query Match: 17.88% Indels: 73
DB: 4 Gaps: 16

US-10-081-864-21 (1-707) x US-09-252-991A-26899 (1-247)

```
QY 8 CTTCCCTGTCAGCAGACCATGCCCTTCAGGACCA-----CCATCGAGGGCA 55
DB 3 ProSerCysSerPro--ProSerProCysSerProLeuArgArgTrpProSerProThr 21
QY 56 CCGTGAACGGCCACTACTCAAGTGCACCGCGAGGGCG----- 94
DB 22 ProAlaThrAlaAsnThrAlaTrpSerProAlaSerProIleArgSerProAlaSerThr 41
QY 95 -----AGGCAACACCCCTCGAGGGCACCCAGAGATGAAGATCGAGGTGATCGAGG 145
DB 42 ThrCysTrpArgCysSerProSer--AlaSerGlyProProSerAsnArgGluProArg 60
QY 146 GCGGCCCTTCCTGCTCCCTTCACATCC---TGTCACCTCCTGCATGACGGCTCCA 202
DB 61 ValSerProCysProAlaTrpSerSerAlaProCysTrpSerAlaGlyCysSerAlaSer 80
QY 203 AGGCTTTCATCA-----AGTACGTGTCGGGCATCCCGACTACTTCAAGCAGTCCCTCC 256
DB 81 ArgAlaCysAsnCysArgSerTrpIleProAlaSerProProArgCys-----Ser 97
QY 257 CCGAGGCTTCACTGGAGCGCACCACCATCTACGAGGACGGCGCTTCCTGACCGCCC 316
DB 98 ProSerAlaSerAlaTrpProSerProCysAlaArgCysArgTrpArgTrpProPro 117
QY 317 ACCAGGACACTCTCGAGCGGCGACTGCTGCTGTACAGGTGAAGATCCTGGGCAACA 376
DB 118 ProArgCysSerProTrp-----Pro 124
QY 377 ACTTCCCGCGAGCGGCCCGCTGATGCAGAACAGCGCGCCCTGGG----- 424
DB 125 ThrGlySerProThrAlaTrpSerCysArgThrSerProAlaProGlyCysThrArgSer 144
QY 425 -----AGCCTTCACCGAGATCGT-----ACGAGGTGGAGCGGCTGCTGC 466
DB 145 ValSerSerProProProProArgCysThrProAlaThrArgTrp-----Cys 161
QY 467 GCGGCC-----AGTCCCTGATGGCCCTGGAGTGC 496
DB 162 AlaSerTrpArgArgProArgProTrpTrpTyrValSerProGlyTrpProArgArgAla 181
QY 497 CCGGCGGTGCCACCTGACCTGCACCTGCACACCATCTACCGCTCCAGAGACCGCCT 556
DB 182 ProGly-----SerGlyCysTrpProAlaGluProProArgLeuPro--- 195
QY 557 CCGCCCTGAAGATGCCGGCTTCCACTTCGAGGACCCAGCATCGAGATCCTGGAGGAG 616
DB 196 -----GlnArgLeuIleSerThrArgArgProTrpCysThrArgArg---ArgPro 211
QY 617 TGG-----AGAAGGCGCAAGTGTCTACAGCAGTACG 646
DB 212 TrpValCysAlaSerIleAlaArgAlaSerGlyArgSerSerThr 226
```

RESULT 11

US-09-472-065A-5
Sequence 5, Application US/09472065A
Patent No. 6638732

GENERAL INFORMATION:
APPLICANT: Evans, Krista

TITLE OF INVENTION: Mutants of Green Fluorescent Protein

FILE REFERENCE: 0942.4020002

CURRENT APPLICATION NUMBER: US/09/472,065A

CURRENT FILING DATE: 1999-12-23

PRIOR APPLICATION NUMBER: US 09/970,762

PRIOR FILING DATE: 1997-11-14

PRIOR APPLICATION NUMBER: US 60/030,935

PRIOR FILING DATE: 1996-11-15

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria, A1 mutant

US-09-472-065A-5

Alignment Scores:

Pred. No.: 1,03e-10 Length: 238
Score: 228.50 Matches: 60
Percent Similarity: 47.47% Conservative: 43
Best Local Similarity: 27.65% Mismatches: 77
Query Match: 17.58% Indels: 38
DB: 4 Gaps: 7

US-10-081-864-21 (1-707) x US-09-472-065A-5 (1-238)

```
QY 13 CTGTCGACGAGACCATGCCCTTCAGGACCATCTCAGGAGCGACCGTGAACGGCCACTAC 72
DB 7 LeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsnGlyHisLys 26
QY 73 TTCAAGTGCACCGCAGCGGAGCGGCAACCCCTCGAGGCGCACCCAGGAGATGAAGATC 132
DB 27 PheSerValSerGlyGluGlyGluGlyAspAlaThrTrpGlyLysLeuThrLeuLysPhe 46
QY 133 GAGGTGATCGAGGGCGCCCTGCTCCCTTCACATCTCCCTGTCACATCTGTCACCTCCTGATG 192
DB 47 ---IleCysThrThrGlyLysLeuProValProTrpProThrLeuValThrThrcysAla 65
QY 193 TAGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCATCCCC-----GACTACTTCAAG 246
DB 66 TyrGlyValGlnCysPheSerArgTrpProAspHisMetLysGlnHisAspPheLys 85
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGCACCATCTACGAGGACGGCGCTTC 306
DB 86 SerAlaMetProGluGlyTyrValGlnGluArgThrIlePhePheLysAspGlyAsn 105
QY 307 CTGACCGCCACCGAGCACCTCCCTGACCGGCGACTCCTGCTGCTGCTACAGGTGAAGATC 366
DB 106 TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 125
QY 367 CTGGGCAACAACCTCCCGCGAGCGGCCCTGATGCAGAACAAAG----- 411
DB 126 LysGlyIleAspPheLysGluaspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyr 145
QY 412 -----GCCGCGCTGGGAGCCCTCCACCGAGATCGTGATC 447
DB 146 AsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 165
QY 448 GAGGTG-----GACGCGCTGCTGGCGCG----- 470
DB 166 LysIleArgHisAsnIleGluaspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn 185
QY 471 ---CCAGTCCCTGATGGCCCTGAGTGCCTCGCGGGTGCACCTGACCTGCCACCTGCA 527
DB 186 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 198
QY 528 CACCACTACCGCTCCAGAGACCGCGCTCCGCG---CCTGAGATGCCCGG 575
DB 199 HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg 215
```

RESULT 12

US-09-417-197-121

Sequence 121, Application US/09417197

Patent No. 6518021

GENERAL INFORMATION:

APPLICANT: Ole THASTRUP, et al.

TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An

CURRENT FILING DATE: 1999-12-23

PRIOR APPLICATION NUMBER: On A Cellular Response

PRIOR FILING DATE: 3759-0110P

CURRENT APPLICATION NUMBER: US/09/417,197

; CURRENT FILING DATE: 1999-10-07
 ; NUMBER OF SEQ ID NOS: 143
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 121
 ; LENGTH: 997
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: EGFP-lkappaB-kinase fusion
 ; US-09-417-197-121

Alignment Scores:
 Pred. No.: 1,5e-10 Length: 997
 Score: 228.50 Matches: 65
 Percent Similarity: 40.98%
 Best Local Similarity: 24.44%
 Indels: 17.58%
 Gaps: 4

US-10-081-864-21 (1-707) x US-09-417-197-121 (1-997)

QY	13	CTGTGACGAGACATCGCTTCCAGGACCATGAGGGGACCGTGAACGGCCACTAC	72
Db	8	LeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsnGlyHisLys	27
QY	73	TTCAGTGCACCGCAGGCGAGGCGAACCCCTCGAGGGGACCCAGGAGATGAATC	132
Db	28	PheSerValSerGlyGluGlyValGluYAspAlaThrTyrGlyLysLeuThrLysPhe	47
QY	133	GAGGTGATCGAGGCGGCGCCCTCGCTTCCATCTCCATCTGCCACCTCGCTCGATG	192
Db	48	--ileCysThrThrGlyLysLeuProValProTyrProThrLeuValThrThrLeuThr	66
QY	193	TACGCTCCAGGCTTCATCAGTACGTGTCCGGCATCCCC-----GACTACTTCAAG	246
Db	67	TyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisaspPheLys	86
QY	247	CAGTCCCTCCCGAGGGCTTCACCTGGGAGGCGACACACCTACGAGGAGCGCGCTTC	306
Db	87	SerAlaMetProGluGlyTyrValGlnGluArgThrIlePhePheLysaspGlyAsn	106
QY	307	CTGACCGCCACACGAGCACCTCCCTCGAGCGGACCTGCTGTGTACAGGTGAATC	366
Db	107	TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu	126
QY	367	CTGGCAACACTTCCCGCGCGCGCGCTGTGTCAGAACAAAG-----411	
Db	127	LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyr	146
QY	412	-----GCGGCGCTGGGAGCCCTCCACCGAGATCGGTAC	447
Db	147	AsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe	166
QY	448	GAGGTG-----GAGCGGTGCTCGCGG-----470	
Db	167	LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn	186
QY	471	---CCAGTCCCTGATGGCTCGGAGTCCCGCGGCGCGCCACCTGACCTGCCACCTGCA	527
Db	187	ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn	199
QY	528	CACCACTTACCGCTCCAGAGACC-----551	
Db	200	HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMet	219
QY	551	-----551	
Db	220	ValLeuLeuGluPheValThrAlaAlaGlyLeThrLeuGlyMetAspGluLeuTyrLys	239
QY	552	-----CGCTCCGCTTGAAGATGCCGGGTTCACCTTCGAGGACACCGCATCGA	602
Db	240	SerGlyLeuArgSerArgAlaGlnAlaSerAsnSerThrMetGluArgProProGlyLeu	259

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:36:21 ; Search time 44.9213 Seconds
(without alignments)
9873.883 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300
Sequence: 1 gggatcgctccctgctgac.....ctggcccaactgaagctt 707

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 2582470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_29072004_150508_25599/app.query.fasta_1.1742
-DB=Published Applications AA -QMT=fastan -SUFFIX=rapp -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -LOCAL=LOCAL -OUTFMT=ptc -NORV=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USPR=US10081864 @CGN 1 1 21 @runat_29072004_150508_25599
-NCFU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -GSPBLOCK=100 -GSAPEXT=0.5
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=7
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description

ALIGNMENTS

RESULT 1
US-10-006-922-40
; Sequence 40, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556

Handwritten signature and initials.

1	1257	96.7	231	13	US-10-006-922-40	Sequence 40, Appl
2	1257	96.7	231	14	US-10-081-864-22	Sequence 22, Appl
3	1251	96.2	232	14	US-10-081-864-24	Sequence 24, Appl
4	1226	94.3	232	13	US-10-006-922-42	Sequence 42, Appl
5	1219	93.8	232	14	US-10-155-809-2	Sequence 2, Appl
6	1219	93.8	232	15	US-10-370-570-59	Sequence 59, Appl
7	1218	93.7	232	14	US-10-155-809-4	Sequence 4, Appl
8	1216	93.5	232	13	US-10-006-922-14	Sequence 14, Appl
9	1216	93.5	232	14	US-10-081-864-10	Sequence 10, Appl
10	1213	93.3	232	16	US-10-724-178-18	Sequence 18, Appl
11	1206	92.8	232	14	US-10-155-809-6	Sequence 6, Appl
12	1045	80.4	197	16	US-10-724-178-1047	Sequence 1047, Ap
13	961	73.9	185	16	US-10-724-178-1065	Sequence 1065, Ap
14	868	66.8	168	16	US-10-724-178-1061	Sequence 1061, Ap
15	785	60.4	152	16	US-10-724-178-1053	Sequence 1053, Ap
16	780.5	60.0	227	14	US-10-155-809-16	Sequence 16, Appl
17	770.5	59.3	226	9	US-09-976-673-10	Sequence 10, Appl
18	770.5	59.3	227	14	US-10-155-809-14	Sequence 14, Appl
19	769.5	59.2	226	9	US-09-976-673-8	Sequence 8, Appl
20	769.5	59.2	226	9	US-09-976-673-16	Sequence 16, Appl
21	769.5	59.2	226	9	US-09-976-673-18	Sequence 18, Appl
22	768.5	59.1	227	9	US-09-976-673-24	Sequence 24, Appl
23	767.5	59.0	227	9	US-09-976-673-6	Sequence 6, Appl
24	766.5	59.0	226	9	US-09-976-673-26	Sequence 26, Appl
25	765.5	58.9	227	9	US-09-976-673-12	Sequence 12, Appl
26	762.5	58.7	227	9	US-09-976-673-2	Sequence 2, Appl
27	762.5	58.7	227	9	US-09-976-673-14	Sequence 14, Appl
28	757.5	58.3	227	14	US-10-155-809-10	Sequence 10, Appl
29	756.5	58.2	227	14	US-10-155-809-12	Sequence 12, Appl
30	753.5	58.0	227	9	US-09-976-673-4	Sequence 4, Appl
31	742	57.1	215	14	US-10-155-809-8	Sequence 8, Appl
32	718	55.2	135	16	US-10-724-178-1051	Sequence 1051, Ap
33	707	54.4	148	15	US-10-370-570-57	Sequence 57, Appl
34	643	49.5	121	16	US-10-724-178-1059	Sequence 1059, Ap
35	637	49.0	228	15	US-10-370-570-60	Sequence 60, Appl
36	570	43.8	111	16	US-10-724-178-1057	Sequence 1057, Ap
37	567.5	43.7	225	16	US-10-423-688A-41	Sequence 41, Appl
38	553	42.5	225	14	US-10-244-779-2	Sequence 2, Appl
39	551.5	42.4	225	15	US-10-442-148A-7	Sequence 7, Appl
40	551.5	42.4	227	15	US-10-442-148A-7	Sequence 7, Appl
41	551.5	42.4	227	15	US-10-370-570-62	Sequence 62, Appl
42	551.5	42.4	236	16	US-10-314-936-4	Sequence 4, Appl
43	550.5	42.3	239	15	US-10-442-148A-8	Sequence 8, Appl
44	550.5	42.3	225	14	US-10-421-258-8	Sequence 8, Appl
45	550.5	42.3	226	16	US-10-724-178-16	Sequence 16, Appl
46	550.5	42.3	226	14	US-10-121-258-6	Sequence 6, Appl

```
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Anemonia sulcata
US-10-006-922-40

Alignment Scores:
Pred. No.: 1.41e-76 Length: 231
Score: 1257.00 Matches: 231
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.69% Indels: 0
DB: 13 Gaps: 0

US-10-081-864-21 (1-707) x US-10-006-922-40 (1-231)
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGCACCGTGAACGGC 66
DB 1 AlaserLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGAAGCCCTTCGAGGCGACCCAGGAGATG 126
DB 21 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTCCATCTCCACCTCC 186
DB 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGCTCGGCGCATCCCGACTACTTCAAG 246
DB 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCTTCACTGGAGCGGACCAACCCCTCGAGGCGACCCAGGAGATG 306
DB 81 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTCGAGCGGCGCCCTTCGCTTCGCTTCACATCTCCACCTCC 366
DB 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGCAACAACCTTCCCGCGAGATGCCCGGCTTCACCTTCGAGGCGACCCAGGCGATC 606
DB 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 200
QY 607 CTGAGAGGTGGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCGCGCGCGCGCTAC 666
DB 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyAspTyr 220
QY 667 TCGACCGCGCGCGCTCCAAAGCTGGGCGCAAC 699
DB 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231

RESULT 2
US-10-081-864-22
; Sequence 22, Application US/10081864
; Publication No. US20030022287A1

; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-22

Alignment Scores:
Pred. No.: 1.41e-76 Length: 231
Score: 1257.00 Matches: 231
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.69% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-081-864-22 (1-231)
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGCACCGTGAACGGC 66
DB 1 AlaserLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGAAGCCCTTCGAGGCGACCCAGGAGATG 126
DB 21 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTCCATCTCCACCTCC 186
DB 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGCTCGGCGCATCCCGACTACTTCAAG 246
DB 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCTTCACTGGAGCGGACCAACCCCTCGAGGCGACCCAGGAGATG 306
DB 81 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTCGAGCGGCGCCCTTCGCTTCGCTTCACATCTCCACCTCC 366
DB 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGCAACAACCTTCCCGCGAGATGCCCGGCTTCACCTTCGAGGCGACCCAGGCGATC 606
DB 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 200
QY 607 CTGAGAGGTGGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCGCGCGCGCTAC 666
DB 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyAspTyr 220
QY 667 TCGACCGCGCGCGCTCCAAAGCTGGGCGCAAC 699
DB 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231
```


Db 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluLe 200
QY 607 CTGGAGGAGTGGAGAGCGGAGTCTACAGAGCAGTACGAGCGCGCGTGGCGCGCTAC 666
Db 201 LeuGluGluValGluLysGlyLysCysIleLysGlnTyrGluAlaValGlyArgTyr 220
QY 667 TCGCAGCGCGCGCGCTCCAAAGTGGCGCACAC 699
Db 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231
RESULT 3
US-10-081-864-24
; Sequence 24, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-24
Alignment Scores:
Pred. No.: 232
Score: 1251.00
Length: 3.57e-76
Matches: 230
Conservative: 0
Percent Similarity: 99.57%
Best Local Similarity: 99.57%
Mismatch: 1
Indels: 0
Query Match: 96.23%
DB: 14
Gaps: 0
US-10-081-864-21 (1-707) x US-10-081-864-24 (1-232)
QY 7 GCCTCCTGCTGACCGAGACCATGCGCTTCAGAGCACCATCGAGGCGCCGCTGACGCGC 66
Db 2 AlaSerLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTCGAGGCGCACCGGAGATG 126
Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 41
QY 127 ARGATCGAGTGTACGAGGCGCGCGCTGCTGCTTCCACATCTGCTCCACTCC 186
Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheIleLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGGCGCACCATCCTACGAGGCGCGGCTTC 306
Db 82 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCGCCACAGGACACTCCCTCGAGCGGGGACTCCCTGGTGTGTACAGGTGAAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121

QY 367 CTGGGCAACAACCTTCCCGCGCGAGCGCGCTGATGCAGAAACAAGCGCGCGCTGGGAG 426
Db 122 LeuGlyAsnAspPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCTCTCACCGAGATCGTGTACGAGGTGGACGGCGTCTCGCGGCGCAGTCCCTGTGATGCGC 486
Db 142 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerSerMetAla 161
QY 487 CTGGAGTGGCGCGCGTGGCGCGCGCTGACCTGCACCTGCACACCTACCGCTCCAG 546
Db 162 LeuGluCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCGCGCTCCCGCTCGAAGATGCCCGCGCTTCCACTTCGAGGACCGCATCGAGATC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluLe 201
QY 607 CTGGAGGAGTGGAGAGCGCAAGTGTCTACAGCAGTACGAGGCGCGCGCTGGGCGCGCTAC 666
Db 202 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TCGCAGCGCGCGCGCTCCAAAGTGGCGCACAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232
RESULT 4
US-10-006-922-42
; Sequence 42, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Anemonia sulcata
US-10-006-922-42
Alignment Scores:
Pred. No.: 1.73e-74
Score: 1226.00
Length: 232
Matches: 223
Conservative: 4
Percent Similarity: 98.27%
Best Local Similarity: 96.54%
Mismatch: 4
Query Match: 94.31%
Indels: 0
DB: 13
Gaps: 0
US-10-081-864-21 (1-707) x US-10-006-922-42 (1-232)

QY 7 GCCTCCTGCTGACCGAGACCATGCGCTTCAGGACCATCGAGGCGCGCTGACGCGC 66
Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGCACCCGAGGATG 126
 Db 22 HistyrPheLysCysThrGlyValGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGGTGATCGAGGGCGGGCCCTGCGCTTCCGCTTCCACATCCTGTCCACCTCC 186
 Db 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCAAAGCCCTTCAAGTACGTGTCGGCATCCCGGCTACTTCAAG 246
 Db 62 CysMetTyrGlySerLysAlaPheileLysTyrValSerGlyileProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTCACTGGGAGCGGACCACTACGAGGACGGCGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACCGAGGACCTCCCTCGAGCGGACCACTACGAGGACGGCGGCTTC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 367 CTGGGCAACAACTTCCCGCGGCGGCTGATGAGCAAGCAAGCGGCGGCTGGAG 426
 Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTTCCACCGAGTGTGTAGAGGTGAGCGGTGCTGCGGCGGCGGCTGATGGCC 486
 Db 142 ProAlaThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 142 ProAlaThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGAGTGGAGAGGGCAAGTGTACAGCAGTACGAGGCGGCGGCTGAGGCGGCTAC 566
 Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGCAGCGCCCGCTCCAAAGCTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 5

US-10-155-809-2
 ; Sequence 2, Application US/10155809
 ; Publication No. US20030092884A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey
 ; APPLICANT: Lukyanov, Konstantin
 ; APPLICANT: Chudakov, Dmitry
 ; TITLE OF INVENTION: Kindling Fluorescent Proteins and
 ; FILE REFERENCE: CLON-073
 ; CURRENT APPLICATION NUMBER: US/10/155,809
 ; CURRENT FILING DATE: 2002-05-24
 ; PRIOR APPLICATION NUMBER: 60/293,752
 ; PRIOR FILING DATE: 2001-05-25
 ; PRIOR APPLICATION NUMBER: 60/329,176
 ; PRIOR FILING DATE: 2001-10-11
 ; PRIOR APPLICATION NUMBER: 10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 09/210,330
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/976,673
 ; PRIOR FILING DATE: 2001-10-12
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 232
 ; TYPE: PRT
 ; ORGANISM: anthozoa

US-10-155-809-2

Alignment Scores:

Pred. No.: 5,11e-74 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservative: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-155-809-2 (1-232)

QY 7 GCCTCCCTCGTGCACGAGACCATGCTTACGAGGACCACTCGAGGGCACCGTCAAGGC 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGCACCCGAGGATG 126
 Db 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGGTGATCGAGGGCGGCGGCTGCGCTTCCACATCCTGTCCACCTCC 186
 Db 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCAAAGCCCTTCAAGTACGTGTCGGCATCCCGGCTACTTCAAG 246
 Db 62 CysMetTyrGlySerLysThrPheileLysTyrValSerGlyileProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTCACTGGGAGCGGACCACTACGAGGACGGCGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACCGAGGACCACTCCCTCGAGCGGCGGCTGCTGAGTACAGGTCAAGTCAAG 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysile 121
 QY 367 CTGGCAACAACTTCCCGCGGCGGCTGATGAGCAAGCAAGCGGCGGCTGGAG 426
 Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTTCCACCGAGTGTGTAGAGGTGAGCGGTGCTGCGGCGGCGGCTGATGGCC 486
 Db 142 ProAlaThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGAGTGGAGAGGGCAAGTGTACAGCAGTACGAGGCGGCGGCTGAGGCGGCTAC 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCGCGCTCCCGCTGAGAGTCCCGGCTTCCACTTCGAGGACCACTCGAGGACGGCGGCTAC 606
 Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgileGluile 201
 QY 607 CTGAGAGGTGGAGAGGGCAAGTGTACAAAGCAGTACGAGGCGGCGGCTGGGCGGCTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGCAGCGCCCGCTCCAAAGCTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 6

US-10-370-570-59
 ; Sequence 59, Application US/10370570
 ; Publication No. US20030219717A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DAHL, Soren Weis et al.
 ; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS
 ; FILE REFERENCE: 3759-0130P
 ; CURRENT APPLICATION NUMBER: US/10/370,570
 ; CURRENT FILING DATE: 2003-02-24
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 59

FILE REFERENCE: CLON-073
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: 60/293,752
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/329,176
PRIOR FILING DATE: 2001-10-11
PRIOR APPLICATION NUMBER: 10/006,922
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 09/210,330
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/976,673
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 232
TYPE: PRT
ORGANISM: anthozoa
US-10-155-809-4

Alignment Scores:

Pred. No.: 5,11e-74 Length: 232
Score: 1219.00 Matches: 221
Percent Similarity: 97.84% Conservative: 5
Best Local Similarity: 95.67% Mismatches: 5
Query Match: 93.77% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-21 (1-707) x US-10-370-570-59 (1-232)

QY 7 GCCTCCCTGTGACCGAGACCATCGCTTCAGGACCAACATCGAGGCGACCGTGAACGCG 66
Db 2 AlaserPheLeuLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAGTCCAGCGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGG 126
Db 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 42 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 82 GlnSerPheProGluGlyPheThrTyrGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCG 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAACTTCCCG 426
Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrGlu 141
QY 427 CCCTCCACCGAGTCTGTACGAGGTGAGCGGCGGTGTGCGCGCGCGCGCGCGCGCGCG 486
Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGGAGTGTCCCG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCG 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGGAGGAGGTGGAGAGGCGCAAGTGTCTACAGCAGTACGAGCGCGCGCGCGCGCG 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
QY 667 TCGGACCG 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 7

US-10-155-809-4
Sequence 4, Application US/10155809
Publication No. US20030092884A1
GENERAL INFORMATION:
APPLICANT: Lukyanov, Sergey
APPLICANT: Lukyanov, Konstantin
APPLICANT: Chudakov, Dmitry
TITLE OF INVENTION: Kindling Fluorescent Proteins and
METHODS FOR THEIR USE

US-10-081-864-21 (1-707) x US-10-155-809-4 (1-232)
QY 7 GCCTCCCTGTGACCGAGACCATCGCTTCAGGACCAACATCGAGGCGACCGTGAACGCG 66
Db 2 AlaserPheLeuLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAGTCCAGCGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGG 126
Db 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 42 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 82 GlnSerPheProGluGlyPheThrTyrGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCG 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAACTTCCCG 426
Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrGlu 141
QY 427 CCCTCCACCGAGTCTGTACGAGGTGAGCGGCGGTGTGCGCGCGCGCGCGCGCGCGCG 486
Db 142 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGGAGTGTCCCG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCG 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGGAGGAGGTGGAGAGGCGCAAGTGTCTACAGCAGTACGAGCGCGCGCGCGCGCG 666
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
 QY 667 TCGGAGCGCGCCCTCCAGCTGGCGGCACAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisasn 232

RESULT 8

US-10-006-922-14
 ; Sequence 14, Application US/10006922
 ; Publication No. US20020197676A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey A
 ; APPLICANT: Fradkov, Arcady F.
 ; APPLICANT: Labas, Yulii A.
 ; APPLICANT: Matz, Mikhail V.
 ; APPLICANT: Tersikh, Alexey
 ; TITLE OF INVENTION: Methods for Using the Same
 ; FILE REFERENCE: CLON-035CIP
 ; CURRENT APPLICATION NUMBER: US/10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 09/120,330
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/457,898
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,144
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,477
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/457,556
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/444,338
 ; PRIOR FILING DATE: 1999-11-19
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 232
 ; TYPE: PRT
 ; ORGANISM: Anemonia sulcata
 US-10-006-922-14

Alignment Scores:
 Pred. No.: 8.14e-74 Length: 232
 Score: 1216.00 Matches: 220
 Percent Similarity: 97.84% Conservative: 6
 Best Local Similarity: 95.24% Mismatches: 5
 Query Match: 93.54% Indels: 0
 DB: 13 Gaps: 0

US-10-081-864-21 (1-707) x US-10-006-922-14 (1-232)

QY 7 GCCTCCTGCTGACCGAGACCATGCTTCAGGACACCATCGAGGCAACCTGTAACGGC 66
 Db 2 AlaserPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 126
 Db 22 HistyPheLysCysThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 41
 QY 127 AGATCGAGGTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 186
 Db 42 LysIleGluValIleGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 61
 QY 187 TGCATGTACGGTCCAAAGCGCTTCAAGTACGAGTGTCCGGCATCCCGCATCTTCAAG 246
 Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGGTTACCTGGGAGGCGCACACCATCTACGAGGCGGGGGTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACAGACACTCCCTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366

Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
 QY 367 CTGGGCAACAACCTTCCCG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCCACCGAGATCGTGTACGAGTGTGACGCGGTGCTGCGCGCGCGCGCGCGCGCG 486
 Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGTGCCTCCCG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCG 606
 Db 182 LysProAlaAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
 QY 607 CTGAGGAGGTGGAGAGGCAAGTGTACAGCAGTACGAGGCGGCGGCGGCGGCGGCTAC 666
 Db 202 MetGluGluValGlyGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
 QY 667 TCGGAGCG 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisasn 232

RESULT 9

US-10-081-864-10
 ; Sequence 10, Application US/10081864
 ; Publication No. US20030022287A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey
 ; APPLICANT: Yankov, Konstantin
 ; APPLICANT: Yankov, Yuriy
 ; APPLICANT: Savitsky, Alexandr
 ; APPLICANT: Fradkov, Arcady
 ; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
 ; FILE REFERENCE: CLON-067
 ; CURRENT APPLICATION NUMBER: US/10/081,864
 ; CURRENT FILING DATE: 2002-06-19
 ; PRIOR APPLICATION NUMBER: 10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 60/270,983
 ; PRIOR FILING DATE: 2001-02-21
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 232
 ; TYPE: PRT
 ; ORGANISM: Anemonia sulcata
 US-10-081-864-10

Alignment Scores:
 Pred. No.: 8.14e-74 Length: 232
 Score: 1216.00 Matches: 220
 Percent Similarity: 97.84% Conservative: 6
 Best Local Similarity: 95.24% Mismatches: 5
 Query Match: 93.54% Indels: 0
 DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-081-864-10 (1-232)

QY 7 GCCTCCTGCTGACCGAGACCATGCTTCAGGACACCATCGAGGCAACCTGTAACGGC 66
 Db 2 AlaserPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 126
 Db 22 HistyPheLysCysThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 41
 QY 127 AGATCGAGGTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 186

Db 42 LysleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACACTCTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyLeuProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGGCGACACCACTACAGAGACGCGGCTTC 306
Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCACACAGGACACCTCCCTGGACGGGACTGCTGTGTACAGGTGAAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValIle 121
QY 367 CTGGGCAACAATCTCCCGCGGACGGCCCGTGTATGACAGAACAAAGCGCGGCTTC 426
Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCCCTCCACCGAGATCGTGTACGAGGTGGAGCGGTGTGCGCGGCGGCTTCGAGGC 486
Db 142 ProAlaThrGluIleValTyrGluValAspGlyValleuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGATGCGCGCGGTGCGACCTGACCTGCGACCTGACACCACTACCGCTCAAG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrTyrArgSerLys 181
QY 547 AAGCGCGCTCCGCGCTGAAGATCCCGGCTTCACCTTCAGGACACCACTGAGATC 606
Db 182 LysProAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGAGAGGTGGAGAGGCAAGTGTACAAGCAGTACAGCGCGCGGCTTCGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGCAGCGCGCGCTCCAGAGTGGGCGCACAAAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 10

US-10-724-178-18
; Sequence 18, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odysey Thera, Inc.
; APPLICANT: Micknick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 232
; TYPE: PPT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: "KFP1" a mutated variant of kindling fluorescent protein asCP fr
; OTHER INFORMATION: m A. sulcata
US-10-724-178-18

Alignment Scores:

Pred. No.:	1,38-73	Length:	232
Score:	1213.00	Matches:	221
Percent Similarity:	96.97%	Conservative:	3
Best Local Similarity:	93.67%	Mismatches:	7
Query Match:	93.31%	Indels:	0
DB:	16	Gaps:	0

US-10-081-864-21 (1-707) x US-10-724-178-18 (1-232)
QY 7 GCCTCCCTGCTGACCGAGACCATGCTTCAGGACCAACCATCGAGGCAACCGTGAACGGC 66
Db 2 AlaSerLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGAACCCCTCGAGGGCAACCCGAGAGATG 126
Db 22 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGGCGGCGGCTTCGCTTCGCTTCACATCTCTGCTCCACCTCC 186
Db 42 LysleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAGTAGCTGTCCGGCATCCCCGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyLeuProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCTTCAGCTGGAGCGGACCACTACAGGACGCGGCTTC 306
Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCACACAGGACACCTCCCTGGACGGGACTGCTGTGTACAGGTGAAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValIle 121
QY 367 CTGGGCAACAATCTCCCGCGGACGGCCCGTGTATGACAGAACAAAGCGCGGCTTC 426
Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 141
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGGAGCGGTGTGCGCGGCGGCTTCGCTGATGGCC 486
Db 142 ProGlyThrGluIleValTyrGluValAspGlyValleuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGATGCGCGCGGTGCGACCTGACCTGCGACCTGACACCACTACCGCTCAAG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrTyrArgSerLys 181
QY 547 AAGCGCGCTCCGCGCTGAAGATGCGCGCTTCACCTTCAGGACCAACCGCTCAAG 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGAGAGGTGGAGAGGCAAGTGTACAAGCAGTACAGCGCGCGGCTTCGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGCAGCGCGCGCTCCAGAGTGGGCGCACAAAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 11

US-10-155-809-6
; Sequence 6, Application US/10155809
; Publication No. US20030092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Chudakov, Dmitry
; APPLICANT: Chudakov, Dmitry
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Their Use
; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10/155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12

```

; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 232
; TYPE: PRT
; ORGANISM: anthozoa
US-10-155-809-6

Alignment Scores:
Pred. No.: 3,846-73 Length: 232
Score: 1206.00 Matches: 219
Percent Similarity: 96.97% Conservativity: 5
Best Local Similarity: 94.81% Mismatches: 7
Query Match: 92.77% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-155-809-6 (1-232)
QY 7 GCCTCCCTGTCGACCGAGACCATCCCTTCAGACCAACCATCGAGGCGACCGTGAACGGC 66
Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTCACCGGACGGGACGGGACCGGACCGGACCGGACCGGACCGGACG 126
Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGluGluMet 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCCCTGCGCTTCGCTTCACATCCCTGTCCACCTCC 186
Db 42 LysIleGluValIleGluGlyGlyProLeuProPhePheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGGCTCAAGGCTTCATCAAGTACGTGTCGGCATCCCGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGACGCGACCAACCATCCACGAGGCGCGCTTC 306
Db 82 GlnSerPheProGluGlyLeuThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCCGCCACAGACACTCCCTGGACGGCGACTGCTGTGTACAGGTGACAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAACCTTCCCGCGGACGCGCGCTGATCAGACAGACGCGCGCGCTGGAG 426
Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGCTGCGGCGCGAGTCCCTGATGGC 486
Db 142 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGGAGTCCCGCGCGCTGCGACCTGATCGACCTGACACACCACTACCGCTCCAG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCCGCTGAGATGCGCGCTTCCACTTCAGAGACCAACCGCATCGAGTC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspTyrArgIleGlu 201
QY 607 CTGGAGAGGTGAGAAAGGCGAGTCTCAAGAGAGTACGAGCGCGCGCTGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGGACCGCGCGCTCCCAAGCTGGGCCACAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 12
US-10-724-178-1047
; Sequence 1047, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odysey Thera, Inc.
; APPLICANT: Michnick, Stephen

```

```

; APPLICANT: MacDonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1047
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F2A, with Met added @ position 1
US-10-724-178-1047

Alignment Scores:
Pred. No.: 2,666-62 Length: 197
Score: 1045.00 Matches: 190
Percent Similarity: 97.96% Conservativity: 2
Best Local Similarity: 96.94% Mismatches: 4
Query Match: 80.38% Indels: 0
DB: 16 Gaps: 0

US-10-081-864-21 (1-707) x US-10-724-178-1047 (1-197)
QY 112 GGCACCCAGGAGATCAAGATCGAGGTGATCGAGGCGCGCCCTGCGCTTCGCTTCAC 171
Db 2 GlyThrGlnGluMetLysIleGluValIleGluGlyProLeuProPheAlaPheHis 21
QY 172 ATCTGTCCACTCTCTGATGTACGGCTCCAGGCTTCATCAAGTACGTGTCGGGATC 231
Db 22 IleLeuSerThrSerCysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIle 41
QY 232 CCCGACTACTTCAAGACAGTCCCTCCCGAGGCTTCACCTGGGAGCGGACCAACACCTAC 291
Db 42 ProAspTyrPheLysGlnSerPheProGluGlyPheThrTrpGluArgThrThrTyr 61
QY 292 GAGACGCGCGCTTCCTGACGCGCCACAGACACTCCCTGGACGCGCTGCTGGTG 351
Db 62 GluAspGlyGlyPheLeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuVal 81
QY 352 TACAAGGTGAGATCTCTGGGCAACAACCTCCCGCGGACGCGCGCTGATCGAGACAA 411
Db 82 TyrLysValLysIleLeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLys 101
QY 412 GCGCGCGCTGGGAGCGCTCCACCGAGATCGTGTACGAGGTGAGCGCGCTGCGCGCG 471
Db 102 ValGlyArgTrpGluProGlyThrGluIleValTyrGluValAspGlyValLeuArgGly 121
QY 472 CAGTCCCTGATGCGCTGAGTCCCGGCGCTGCGGACCTGACCTGACCTGCGACAC 531
Db 122 GlnSerLeuMetAlaLeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThr 141
QY 532 ACCTTACCGCTCCCAAGAGCGCGCTCCCGCTCAAGATGCGCGCTTCCACTTCGAGGAC 591
Db 142 ThrTyrArgSerLysLysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAsp 161
QY 592 CACCGCATCGAGATCTGAGGAGGTGAGAGGGAAGTGTACAGCAGTACGAGGCG 651
Db 162 HisArgIleGluIleMetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAla 181
QY 652 GCGGTGGCGCGCTACTGCGACGCGCGCGCTCCCAAGCTGGGCCACAC 699
Db 182 AlaValGlyArgTyrCysAspAlaAlaProSerLysLeuGlyHisAsn 197

RESULT 13
US-10-724-178-1065
; Sequence 1065, Application US/10724178
; Publication No. US20040137528A1

```


GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1065
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.: KFP F1F, with position 1 Met removed
US-10-724-178-1065

Alignment Scores:
Pred. No.: 128-56 Length: 185
Score: 961.00 Matches: 176
Percent Similarity: 96.22% Conservative: 2
Best Local Similarity: 95.14% Mismatches: 7
Query Match: 73.92% Indels: 0
DB: 16 Gaps: 0

US-10-081-864-21 (1-707) x US-10-724-178-1065 (1-185)
QY 7 GCCTCCCTGTCAGGAGACCATGCTTCCAGGACCATGCTGAGGAGCCGCTGAGACGGC 66
Db 1 AlaserLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTCACCGGCAAGGCGAGGCGGCAACCCCTCCAGGCGACCCAGGAGATG 126
Db 21 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 186
Db 41 LysileGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAG 246
Db 51 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGAGCGGCGACCACTGCTGCTGCTGCTGCTGCTGCT 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTrpGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTGAGCGGCGACTGCTGCTGCTGCTGCTGCTGCTGCT 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGCAACAATCTCCCGCGAGCGGCGGCTGATGAGAACAGGCGGCGGCTGAGATC 426
Db 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 140
QY 427 CCCTCCACCGAGATCGTGTACGAGTGTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
Db 141 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGGAGTCCCGGCGGTGCGCACTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 546
Db 161 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrThrArgSerLys 180
QY 547 AAGCCCGCTCCGCC 561
Db 181 LysProAlaSerAla 185

RESULT 14

US-10-724-178-1061
; Sequence 1061, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1061
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.: KFP F1E, with position 1 Met removed
US-10-724-178-1061

Alignment Scores:
Pred. No.: 22e-50 Length: 168
Score: 868.00 Matches: 159
Percent Similarity: 95.83% Conservative: 2
Best Local Similarity: 94.64% Mismatches: 7
Query Match: 66.77% Indels: 0
DB: 16 Gaps: 0

US-10-081-864-21 (1-707) x US-10-724-178-1061 (1-168)
QY 7 GCCTCCCTGTCAGGAGACCATGCTTCCAGGACCATGCTGAGGAGCCGCTGAGACGGC 66
Db 1 AlaserLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTCACCGGCAAGGCGAGGCGGCAACCCCTCCAGGCGACCCAGGAGATG 126
Db 21 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 186
Db 41 LysileGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAG 246
Db 61 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGAGCGGCGACCACTGCTGCTGCTGCTGCTGCTGCT 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTrpGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTGAGCGGCGACTGCTGCTGCTGCTGCTGCTGCTGCT 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGCAACAATCTCCCGCGAGCGGCGGCTGATGAGAACAGGCGGCGGCTGAGATC 426
Db 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 140
QY 427 CCCTCCACCGAGATCGTGTACGAGTGTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
Db 141 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGGAGTCCCGGCGGTGCGCACTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 510
Db 161 LeuLysCysProGlyGlyArgHis 168

RESULT 15
US-10-724-178-1053


```
; Sequence 1053, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: Patent version 3.0
; SEQ ID NO 1053
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KPP F1C, with position 1 Met removed
US-10-724-178-1053
```

```
Alignment Scores:
Pred. No.:      8.52e-45      Length:      152
Score:          785.00      Matches:     144
Percent Similarity: 95.39%      Conservative: 1
Best Local Similarity: 94.74%      Mismatches: 7
Query Match:      60.38%      Indels: 0
DB:              16      Gaps: 0
```

US-10-081-864-21 (1-707) x US-10-724-178-1053 (1-152)

```
QY 7 GCCTCCCTGCTGACGAGACATGCCCTTCAGGACACCATCGAGGCGACCGTGAACGGC 66
Db 1 AlaSerLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTGCACCGGAGGGCGAGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db 21 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGTGATCGAGGGCGGCCCTTCCTCCCTTCACATCCATCTCCACCTCC 186
Db 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAGTAGTGTCTCGGCATCCCGACTACTTCAAG 246
Db 61 CysMetTyrGlySerLysThrPheIleLysLysValSerGlyLeuProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCGCTTCACTGGGAGCGCACCCACCATCTACGAGGCGCGGCTTC 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACAGACACCTCCCTGGAGCGGCACTGCTGGTGTAAGGTGAAGATC 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGGCAACAACCTTCCCGCCGAGCGGCCCTGATGCAGAACAAAGCGCGCGCTGGGAG 426
Db 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 140
QY 427 CCTCCACCGAGATCGTGACAGGTGGACGGCGTG 462
Db 141 ProGlyThrGluIleValTyrGluValAspGlyVal 152
```

Search completed: July 29, 2004, 14:54:52
Job time : 47.9213 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:28:05 ; Search time 15.3141 Seconds

(without alignments)
8881.684 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300

Sequence: 1 ggatccgctcctcctgac.....ctgggccacaactgaagctt 707

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2 1/USPTO.spool_p/US10081864/runat 29072004 150506 25492/app_query.fasta_1.1742
-DB=pir 78 -QMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 @CGN 1 1 33 @runat 29072004 150506 25492 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	257	19.8	3020	2 A43932	mucin 2 precursor,
2	249.5	19.2	464	2 S22697	extensin - Volvox
3	231.5	17.8	539	2 T28770	hypothetical prote
4	229	17.6	620	2 S06733	hydroxyproline-ric
5	222	17.1	580	2 T43481	probable mucin DKF
6	221.5	17.0	238	1 J01514	green-fluorescent
7	221.5	17.0	1188	2 S49915	extensin-like prot
8	218	16.8	760	2 T86387	probable Pto kinas
9	209	16.1	214	2 T09854	proline-rich cell
10	209	16.1	708	2 D96711	hypothetical prote
11	208	16.0	222	2 H96711	hypothetical prote
12	207	15.9	214	2 T10737	extensin-like cell
13	207	15.9	1201	2 G86441	unknown protein [i
14	203.5	15.7	351	2 S50754	hypothetical prote

ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999

C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, S.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the

A:Reference number: A49963; MUID:94132002; PMID:8300571

A:Accession: A49963

A:Molecule type: mRNA

A:Residues: 1-639 <GU>

A:Cross-references: GB:L21998

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr

A:Reference number: A45106; MUID:93016075; PMID:1400449

A:Accession: A45106

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU>

A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; FID:g186396

A:Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 2037-3020 <GU>

A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; FID:g186398

A:Experimental source: colon

A:Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M

J. Clin. Invest. 88, 1005-1013, 1991

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp

A:Reference number: A43932; MUID:91358717; PMID:1885763

A:Accession: A43932

A:Molecule type: DNA

extensin-like prot
hypothetical prote
hydroxyproline-ric
antifreeze glycope
extensin homolog F
hydroxyproline-ric
cell wall glycopro
mucin 5AC (clone J
hydroxyproline-ric
hypothetical prote
extensin homolog T
hypothetical 119.5
hypothetical 119.5
pTB-associated spl
extensin homolog F
extensin-like prot
mucin 7 precursor,
hypothetical prote
proline-rich prote
hypothetical prote
probable cell wall
probable membrane
high molecular mas
tegument protein 2
hypothetical prote
salivary glue prot
pistil extensin-li
Miskott-Aldrich sy
Miskott-Aldrich sy
probable proline-r
mucin - rhesus mac

A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
 A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBI:P:55750)
 R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
 J. Biol. Chem. 264, 6480-6487, 1989
 A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence
 A:Reference number: A33532; MUID:89197956; PMID:2703501
 A:Accession: B33532
 A:Molecule type: mRNA
 A:Residues: 1916-2193 <GU4>
 A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
 A:Experimental source: intestine
 R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Baasbaum, C.B.
 J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481; PMID:1985113
 A:Accession: A61257
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner,
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-
 A:Reference number: P00328; MUID:92198477; PMID:1550588
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
 C:Keywords: Glycoprotein; intestine; tandem repeat
 F:2786-2834/Domain: von Willebrand factor type C repeat homology <WVC>
 Alignment Scores:

Pred. No.:	1.56e-09	Length:	3020
Score:	257.00	Matches:	83
Percent Similarity:	39.93%	Conservative:	24
Best Local Similarity:	30.97%	Mismatches:	93
Query Match:	19.77%	Indels:	68
DB:	2	Gaps:	13

US-10-081-864-21 (1-707) x A43932 (1-3020)

QY 5 CCGCTT-----CCCTGCTGACGAGACCATGCGCTTCAGGACCAACCATCGAGG---- 52
 Db 1495 ProProThrThrThrProSerProProMetThrThrProleThrProProAlaSerThr 1514
 QY 53 -----GCACCGTGAACGGCCACTACT----- 73
 Db 1515 ThrThrLeuProProThrThrThrProSerProProThrThrThrThrThrProPro 1534
 QY 74 ---TCAAGTACCGGACGGGAGGAGGCAACCCCTCGAGGACCAACCGAGATGAAGA 130
 Db 1535 ProThrThrThrProSerProProThrThrThrProleThrProleThrProSerThrThr 1554
 QY 131 TCAGAGTGATCGAGGGGGCCCTGCTGCTTCCATCCATCCCTGCTCCACTCTGCA 190
 Db 1555 ThrLeuProProThrThrThrProSerProProThrThrThrThrThrProProPro 1574
 QY 191 TGTACGGCTCAAGGCTTCATCAAGTACGTGTCGGGATCCCGACT---ACTCAAGC 247
 Db 1575 ThrThrProSerProProThrThrThrProSerProProThrThrThrThrThrThr 1594
 QY 248 AGTCCCTCCCGAGGGGCTTCACCTGG-----AGCGCACCAACCACT 289

Db 1595 ThrProProProThrThrThrThrProSerProProThrThrThrThrThrThrThrThr 1614
 QY 290 ACAGAGACGGCGCTTCCTGACGGCCACCAAGACACCT-----CCCTGGACGGGACT 343
 Db 1615 ThrThrThrProSer-----ProProThrThrThrThrThrThrThrThrThrThr 1632
 QY 344 GCGTGTGTACAAAGGTGAAGATCCTGGGCAACAACTTCCCGCCGACGCGCCCGGTGATGC 403
 Db 1633 -----ThrThrLeuProProThr----- 1638
 QY 404 AGAACHAGCCCGCGCTGGGAGCCCTCCACGAGATCGTGACAGGTGGACGCGTGC 463
 Db 1639 ---ThrThrPro-----SerProProProThrThrThrThrThrThrThrThrThrThr 1654
 QY 464 TCGCGCGCCAGTCCCTGATGGCCCTGGAGTCCCGGGGTGCGCACCTGACCTGCCACC 523
 Db 1655 ThrThrProSerPro-----ProThrThrThrThrThrThrThrThrThrThrThr 1662
 QY 524 TGCACA-----CCACTACCGCTCCAAAGACCGCCCTCCGCTCGAGATGCCCGCT 577
 Db 1663 ThrThrProSerProProleThrThrThrThrThrThrThrThrThrThrThrThrThr 1682
 QY 578 TCACCTTCGAGGACCAACCGCATCGAGATCCTGGAGGAGGTGGAGAGGCAAGTCTACA 637
 Db 1683 SerProleThrThrThrProSerProProThrThrThrThrThrThrThrThrThrThr 1702
 QY 638 AGCAGTACGAGCGCCCGCTGGCGCTACTCGAGCGCG----- 676
 Db 1703 ThrThrProSerProleThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1722
 QY 677 CCGCTTCAAGCTGGCGCAACT 700
 Db 1723 ProProProThrThrMetThrThr 1730
 RESULT 2
 S22697
 A:extensin - Volvox carteri (fragment)
 C:Species: Volvox carteri
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
 C:Accession: S22697; S21006
 R:Erli, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
 EMBO J. 11, 2055-2062, 1992
 A:Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca
 A:Reference number: S22697; MUID:92289669; PMID:1600938
 A:Accession: S22697
 A:Molecule type: mRNA
 A:Residues: 1-464 <HAL>
 A:Cross-references: EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
 C:Keywords: glycoprotein
 Alignment Scores:

Pred. No.:	4.87e-09	Length:	464
Score:	249.50	Matches:	86
Percent Similarity:	37.55%	Conservative:	12
Best Local Similarity:	32.95%	Mismatches:	79
Query Match:	19.19%	Indels:	85
DB:	2	Gaps:	13

US-10-081-864-21 (1-707) x S22697 (1-464)

QY 5 CCGCTTCCCTGCTGACGAGACCATGCGCTTCAGGACCAACCA-----TCGAGG 52
 Db 223 ProProPro-----ProArgValSerThrSerProProProAlaArgValSerSer 240
 QY 53 GCAGGTGAACGGCGCATCTTCAAGTGCACCGCGCAGG----- 91
 Db 241 SerProProAlaThrArgSerProProProArgArgIleThrSerProSerProVal 260
 QY 92 -----GCAGGGCAACCCCTTCGAGGGCACCAGGATGAAGA 130
 Db 261 LeuThrAlaSerProProLeuProLeuThrSerProProProPro----- 276

QY 131 TCAGGTGATCGAGGCGGCCCTTGCCTTCCACATCTCTGCTCCACATCTCTGCA 190
 Db 277 -----ArgValProProSerProProProValAlaSerProProPro 292
 QY 191 TGTACGGCTCCAGGCGCTTCATCAAGTGTCTCCGGATCCCGACTCTCAAGCAGT 250
 Db 293 ---ProProProArgValSerProSerProProProProGlnProValSerSerProPro 311
 QY 251 CCTCCCGGAGGCTTCACCTGGAGCGCACACACCTACAGGAGCGGCGCTTCTGCA 310
 Db 312 ProProProProProProSerProSerProProProProProArgSerSerProPro 331
 QY 311 CCGCCACACAGGACACTCTCTGAGCGGCGACTGCTGTGTACAGGTGAAGATCTCTGG 370
 Db 332 ProProProSerProProPro----- 338
 QY 371 GCAACAACCTTCCCGCGCGAGCGCCCGTGTATGAGAACAAAGCGCGCGCTGGAGCCCT 430
 Db 339 -----ProSerProProPro-----ProArgProSerProSerProPro 351
 QY 431 CCACCGAGATCGGTACGAGGTGGACGGCGTGTGCGCGGCGAGTCCCTGATGGCCCTGG 490
 Db 352 ProProArgSer-----Ser-ProSerProProProProVa 363
 QY 491 AGTG-----CCCGCGGCTCGCCACCTGACCTCCACCTGCACA--- 529
 Db 363 lValSerProProProProProArgAlaSerProProPro-----ProProAlaSerSe 382
 QY 530 -----CCACCTACCGCTCCAAGAACGCCCGCTCCCGCCCTG 564
 Db 382 rProProProProProProProProProProProProProProProProProProAl 402
 QY 565 AAGATGCCCGCTTCACCTCGAGGACCCCGCATCGAGATCTCGAGGAGGTGGAGAG 624
 Db 402 aThrAlaAlaAlaSerProProSerProAlaProSerArgSer-----ArgAl 418
 QY 625 GGCAAGTGTACAAGACAGTACGAGCGCGCTGGCGCTACTCGACGCGCGCCCTCC 684
 Db 418 aGlyGlyProProLeuGlyThrArgProPro-----ProProProPr 432
 QY 685 A 685
 Db 432 O 432

RESULT 3

T28770
 hypothetical protein W03D2.1 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 R;Rohlfing, T.; Wohldmann, P.
 submitted to the EMBL Data Library, June 1997
 A;Description: The sequence of C. elegans cosmid W03D2.
 A;Reference number: Z20519
 A;Accession: T28770
 A;Status: Preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-539 <ROH>
 A;Cross-references: EMBL:AF000298; PIN:AA48255.1; GSPDB:GN00022; CESP:W03D2.1
 A;Experimental source: strain Bristol N2; clone W03D2
 C;Genetics:
 A;Gene: CESP:W03D2.1
 A;Map position: 4
 A;Introns: 40/3; 88/3; 115/3; 146/3; 173/3
 C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Alignment Scores:
 Pred. No.: 7.54e-08 Length: 539
 Score: 231.50 Matches: 87
 Percent Similarity: 38.87% Conservative: 23
 Best Local Similarity: 30.74% Mismatches: 103
 Query Match: 17.81% Indels: 70
 DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x T28770 (1-539)
 QY 5 CCGCTCTCTCTGCTGACGAGACCATGCTTCAAGCAACCATCCAGGGCACCGTGAACG 64
 Db 253 ProProAlaGlySerProProProPro-----ProProLysGlySerProProLeu 270
 QY 65 GCGACTCTCAAGTGCACCGGCAAGCGGAGGCAACCCCTCGAGGGACCCAGGAGA 124
 Db 271 AlaGlySerGlySerProProProProProProAlaAlaGlySerProProProProArg--- 289
 QY 125 TGAAGATCGAGGTGATCGAGGGCGCCCTTCCGCTTCCGCTTCCACATCTCTGCTCCACT 184
 Db 290 -----ThrGlySerProProProProProProProProGlySerProProPro 303
 QY 185 CTGCTCATGT-----ACGGCTCCAAGGCTTCATCAAGTACGTCTCG----- 226
 Db 304 ProProAlaGlyGlySerProProProProProProProProProProProProPro 323
 QY 227 -----GCATCCCGGACTACTTCAAGCAGTCCCTCCCGAGGGTTCACCTGGGAGCGCA 280
 Db 324 ProArgGlySerProProProProProProProProProProProProGlnAla-----GlyGlySer 341
 QY 281 CCACCACTACGAGGA-----CGCGGCTTCC 307
 Db 342 ProProProAla-GlyThrGlySerProProProProProProProProArgGlnAlaPr 361
 QY 308 TGACCGC-----CCACGAGACACCTCCCTGG 334
 Db 361 oGluArgSerProProProProProProProProProProProProProProProPro 381
 QY 335 ACGGCGACTGCTGTGTACAAAGTGAAGATCTCGGCAACAACCTTCCCGCCGACGCGCC 394
 Db 381 oArgGlyGlyProGlyLysSerSerGluGluSerSerGluSer-----ArgGluGlyPr 399
 QY 395 CGGTATCGACAACAGCGCGCGCTCGGA-----GC 427
 Db 399 oArgGlyGlyProArgGlyGlyProArgGlyGlyProArgLysSerSerGluGluSerSe 419
 QY 428 CCTCCACGAGATCGTGTACGAGGTGGACGGCTGCT-----GCGCGCCAGTCC 478
 Db 419 rGluSerArgGluGluProArgGlyProArgSerProProProProProProTh 439
 QY 479 TGATGGCCCTGGAGTG-----CCCGCGGCTGCGCCTGACCTGACCTGCGCACCTGCA 529
 Db 439 rGlySerProProProProProProProProProProProProProProProPro 459
 QY 530 CCACCTACCGTCCAGAGAGCGCGCTCCGCGCTCAAGAT----- 569
 Db 459 rGlyLeuProSerArgGlnLysArgGlnAlaProGluAspArgProProProProTh 479
 QY 570 -GCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATCTCGAGGAGGTGGAGAGGCA 628
 Db 479 oThrGlySerProProProProProProProProProProProProProProPro 497
 QY 629 AGTGCTACAGCAGTACAGGCGCGCTGGCGCGCTACTGGGACCGCCCTCCAGC 688
 Db 497 uSerSerGluSerA-gGluGlyProArgGlyPro-----ArgArgSerProProProTh 515
 QY 689 TGGGCCA 695
 Db 515 ySerPro 517

RESULT 4

S06733
 hydroxyproline-rich glycoprotein precursor - common tobacco
 C;Species: Nicotiana tabacum (common tobacco)
 C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
 C;Accession: S06733
 R;Keller, B.; Lamb, C.J.
 Genes Dev. 3, 1639-1646, 1989
 A;Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene
 A;Reference number: S06733; MUID:90128263; PMID:2612909

A:Accession: S06733
A:Molecule type: DNA
A:Residues: 1-620 <KE>
A:Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein

Alignment Scores:
Pred. No.: 1.1e-07 Length: 620
Score: 229.00 Matches: 81
Percent Similarity: 40.23% Conservative: 22
Best Local Similarity: 31.64% Mismatches: 89
Query Match: 17.62% Indels: 64
DB: 2 Gaps: 11

US-10-081-864-21 (1-707) x S06733 (1-620)

```
QY 5 CCGCTCCCTGCTGACCGAGACCATGCCCTTCAGACCAACCATCGAGG----- 52
Db 207 ProProProThrHisValGlnProThrProSer---ProProSerArgGlyHisGlnPro 225
QY 53 -----GCACCGTGAACGGCCACTTCAAGTGCACCGGCAAGG 91
Db 226 GlnProProThrHisArgHisAlaPro-----ProThrHisArgHisAlaProProThr 243
QY 92 CGAGGGCAACCCCTCGAGG-----GCACCGAGGAGATGAAGATCGAGGTGA 139
Db 244 HisGlnProSerProLeuArgHisLeuProProSerProArgGlnProGlnProPro 263
QY 140 TCGAGGGGG-----GCCCTGCCCTTCGCTTCCACATCC 175
Db 264 ThrTySerProProProProAlaTyAlaGlnSerProGlnProSerProProProPro 283
QY 176 TGTCCACTCTGCTGATGACGGCTTCAAGGCTTCAAGTACGTTCGGCATCCCG 235
Db 284 ProProProProProProProProProProProProProProProProProProPro 303
QY 236 ACTACTTCAAGCAGTCCCTCCCGAGGCTTCACTGGGAGCGCACCACCACTACGAGG 295
Db 304 AlaTySerProProProProProProProProProProProProProProProProPro 321
QY 296 ACGGCGGCTTCTGACCGCCACCGAGACACTCCCTCGGACGGGACTGCTGTGTACA 355
Db 322 ProAlaTySerProProProProProProProProProProProProProProProPro 333
QY 356 AGGTGAAGATCTGGGCAACACTTCCCGCGAGCGCCCGGTGATCGAACAAGCCG 415
Db 334 -----ProProThrTyLeuProLeuProSerProPro 344
QY 416 GCCGCTGGAGCCCTCCACCGAGATCGTGTACGAGGTGGAGCGGCTGCTGGCGCCAGT 475
Db 345 IleTySerProProProPro-----ValTySer 354
QY 476 CCTGTAGTGGCTGGAGTCCCGGGGTGCGCACCTGACTGCTGCACCTGCACCACT 535
Db 355 ProProProPro-----ProSerTySerProProProProProProProProProPro 371
QY 536 ACCGCTCAAGAAGCCGCTCGCCCTGAAGATGCCCGGCTTCCACTTCGAGGACCACC 595
Db 372 ProProProSerProProProProProProProProProProProProProProProPro 391
QY 596 GCA-----TCGAGATCTGAGGAGGTGGAGAGGCAAGTGTCTACAGCAGTACG 646
Db 392 SerProProProProProAlaTySerProProLeuProAlaProProProProProPro 411
QY 647 AGGCGCGCTGGCGGCTACTGCGACCGCCCGCCCTCCAGCTGGGCC 694
Db 412 ProProPro-----ThrTySerProProProProProProProProProProPro 424
```

RESULT 5

T43481
probable mucin DKFZp434C196.1 - human (fragment)
N:Alternate names: protein DKFZp434B0635.1

C:Species: Homo sapiens (man)
C>Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text_change 20-Apr-2000
C:Accession: T43481; T34549; T17264
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z22514
A:Accession: T43481
A:Molecule type: mRNA
A:Residues: 1-580 <AAA>
A:Cross-references: EMBL:AL133561; NID:g6599133; PIDN:CAB63715.1; PID:g6599134
A:Experimental source: adult testis; clone DKFZp434C196
R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A:Reference number: Z21540
A:Accession: T34549
A:Molecule type: mRNA
A:Residues: 262-580 <POU1>
A:Cross-references: EMBL:AL122069; NID:g6102864; PIDN:CAB59245.2; PID:g7018420
A:Experimental source: adult testis; clone DKFZp434B0635
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18723
A:Accession: T17264
A:Molecule type: mRNA
A:Residues: 262-580 <POU2>
A:Cross-references: EMBL:AL117481; NID:g5911958; PIDN:CAB55954.1; PID:g5911959
A:Experimental source: adult testis; clone DKFZp434B061
C:Genetics:
A:Note: DKFZp434C196.1; DKFZp434B0635.1; DKFZp434B061.1

Alignment Scores:
Pred. No.: 3.2e-07 Length: 580
Score: 222.00 Matches: 94
Percent Similarity: 39.55% Conservative: 12
Best Local Similarity: 35.07% Mismatches: 95
Query Match: 17.08% Indels: 68
DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x T43481 (1-580)

```
QY 20 CCGAGACCATGCCCTTCAGGACCAACCATCGAGGCA-----CCGTGA 61
Db 65 ProAlaSerLeuMetArgThrProThrArgAlaSerLeuMetArgThrProProArg 84
QY 62 ACGGCCACTACTTCAAGTGCACCGGACGGGCAACCCCTCGAGGCA---CCC 118
Db 85 AlaSerProThrArgLysProProArgAlaSerProArgThrProSerArgAlaSerPro 104
QY 119 AGGAGATGAAGATCGAGGTGATCGAGGCGGCCCTCGCTTCCCATCATCTGT 178
Db 105 ThrArg-----ArgLeuProArgAlaSerProMetGlySerProHisArgAla--- 120
QY 179 CCACCTCTGTGATGATCGGCTCCAGGCTTCATCAAGTGTCTCGGGATCCCGACT 238
Db 121 ---SerProMetArgThrProProArg-----AlaSerProThr 132
QY 239 ---ACTTCAAGCAGTCCCTCCCGAGGCTTCACTGGGAGCGCACCACTACGAGG 295
Db 133 GlyThrProSerThrAlaSerProThrGlyThrProSerSerAla---SerProThrGly 151
QY 296 ACGGCGGCTTCTGACCGCCACCGAGGACCTCCCTGGAGCGGCGACTGCTGTGTACA 355
Db 152 ThrProProArgAlaSerProThrGlyThrPro----- 163
QY 356 AGGTGAAGATCTGGGCAACACTTCCCGCGAGGCGGCCCTGTATGCAGACAAGG--- 412
Db 164 -----ArgAlaTrpAlaThrArgSerProSerThrAlaSerLeuThrArgThrProSer 181
QY 413 -----CCGCGCGCTGGAGCCCT-----CCACCGAGATCG 442
Db 182 ArgAlaSerLeuThrArgTrpProProArgAlaSerProThrArgThrProProArgGlu 201
QY 443 TGTACGAGGTGAGCGGCGTGTCTGCGGCGCACTCTCTGATGGCCCT---GGAGTGGCCCG 499
```

Db 202 SerProArgMetSer-HisArgAlaSerProThrArgThrProProArgAlaSerProTh 221
QY 500 GCGGTGCCACCTGACCTGCCACCTGCACACACCTACCGCTCCAGAGACCGCGCTCG 559
Db 221 xArgArgProProArgAlaSerProThrArgThrProProArgGluSer-----LeuAr 239
QY 560 CCTGAGATGCCCGCTTCCACTTCGAGGACACCGCATCGAGATCCTGG-AGGAGTGG 618
Db 239 gThrSerHisArgAlaSerProThrArgMetProProArgAlaSerProThrArgArgPr 259
QY 619 GAGAGGCGCAAGTGCCTACAGCAGT-----ACGAGCCCGCTGGGCG 660
Db 259 oProArgAlaSerProThrGlySerProProArgAlaSerProMetThrProProArgAl 279
QY 661 CGTACTGCGACCGCCGCC----- 679
Db 279 aSerProArgThrProProArgAlaSerProThrThrThrProSerArgAlaSerLeuTh 299
QY 680 ----CCTCCAAAGCTGGGCCACA 697
Db 299 rArgThrProSerTrpAlaSer 306

RESULT 6
JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C/Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C/Accession: JS0892; JQ1514, FQ0335; S48693; S51330; S51331
R/Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Friedberg, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A/Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A/Reference number: JQ1514; MUID:92175527; PMID:1347277
A/Accession: JS0692
A/Molecule type: DNA
A/Residues: 1-107, 'S', 109-238 <PRA1>
A/Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663
A/Accession: JQ1514
A/Molecule type: mRNA
A/Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A/Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661
A/Accession: FQ0335
A/Molecule type: protein
A/Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>
R/Incuve, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A/Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A/Reference number: S48693; MUID:94364470; PMID:8082767
A/Accession: S48693
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A/Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384
R/Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A/Reference number: S51330
A/Accession: S51330
A/Molecule type: mRNA
A/Residues: 1-13, 'V', 15-24, 'Q', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R',
A/Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009
A/Experimental source: clone gfp1
A/Accession: S51331
A/Molecule type: mRNA
A/Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q',
A/Cross-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011
A/Experimental source: clone gfp2
R/Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A/Reference number: A65692; PDB:1LGL
A/Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-9
A/Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R/Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996

A/Title: The molecular structure of green fluorescent protein.
A/Reference number: A58953; MUID:98294543; PMID:9631087
A/Contents: annotation: X-ray crystallography, 1.9 angstroms
C/Comment: This protein is excited by the photoprotein aequorin (see PIR:AGJFNV) emitting
C/Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C/Genetics:
A/Gene: GFP
A/Introns: 69/3; 167/3
A/Superfamily: green-fluorescent protein
C/Keywords: chromoprotein; luminescence
F:65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Alignment Scores: 3.45e-07 Length: 238
Pred. No.: 221.50 Matches: 60
Score: 47.00% Conservative: 42
Percent Similarity: 27.65% Mismatches: 78
Best Local Similarity: 17.04% Indels: 38
Query Match: 1 Gaps: 7
DB:

US-10-081-864-21 (1-707) x JQ1514 (1-238)

QY 13 CTGCTGACGAGACCATGCTTCAGGACACCATCGAGGACCGTGAACGGCACTAC 72
Db 7 LeupheThrGlyValValProIleuValGluLeuAspGlyAspValAsnGlyHisLys 26
QY 73 TTCAGTGCACCGGACGAGGCGGCAACCCCTCGAGGCGCACCCAGGAGATGAAGATC 132
Db 27 PheSerValSerGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPhe 46
QY 133 GAGGTGATCGAGGCGGCGGCGGCTCCCTCCCTCCACATCCCTGCTCCACCTCGTCATG 192
Db 47 ---IleCysThrThrGlyLysLeuProValProThrProThrLeuValThrPheSer 65
QY 193 TACGCTCCAAAGCGCTTCATCAAGTACGTGTCGCGCATCCCC-----GACTACTTCAAG 246
Db 66 TyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAspPheLys 85
QY 247 CAGTCCCTCCCGAGGCTTACCTGGGAGGCGGCGGCGGCGGCGGCGGCTTC 306
Db 86 SerAlaMetProGluGlyTyrValGlnGluArgThrIlePheTyrLysAspGlyAsn 105
QY 307 CTGACCGCCACGAGGACCTCCCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 366
Db 106 TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 125
QY 367 CTGGGCAACAACCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 411
Db 126 LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysMetGluTyrAsnTyr 145
QY 412 -----CCCGCGGCTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 447
Db 146 AsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 165
QY 448 GAGGTG-----GACGCGGCTGCTGCCTCGCGG----- 470
Db 166 LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn 185
QY 471 ----CAGTCCCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 527
Db 186 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 198
QY 528 CACCACTACCGCTCCCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 575
Db 199 HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg 215

RESULT 7
S49915
extensin-like protein - maize
C/Species: Zea mays (maize)
C/Date: 05-Mar-1995 #sequence revision 12-May-1995 #text change 29-Oct-1999
C/Accession: S49915

R.Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A:Description: Pex genes: pollen-specific genes with extensin-like domains.

A:Reference number: S49915

A:Accession: S49915

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1188 <RUB>

A:Cross-references: EMBL:234465; NID:5600117; PIDN:CAA84230.1; PID:g600118

Alignment Scores:
Pred. No.: 3,46e-07 Length: 1188
Score: 221.50 Matches: 68
Percent Similarity: 38.59% Conservative: 25
Best Local Similarity: 28.22% Mismatches: 73
Query Match: 17.04% Indels: 75
DB: 2 Gaps: 9

US-10-081-864-21 (1-707) x S49915 (1-1188)

```
QY 8 CCTCCCTGCTGACGAGACCATGGCT-----TCAGGACCAACA----- 46
Db 924 ProProMetThrProLysSerProProValValSerProProThr 943
QY 47 ---TCGAGGGCAGCGTGAACCGCCACTACTTCAAGTCGACCGCAAGGGGAGGCAAC 103
Db 944 VallysSerProProAlaProValSerProProAlaThrProLysSer 963
QY 104 CCTCGAGGGACCCAGAGATGAAGATCGAGGTGTCGAGGGGCGCCCTCGCCCTCG 163
Db 964 ProProAlaProValAsnLeuProProGluProValLysSerProProThr 983
QY 164 CTTCCACATCTGTCCACCTCTCGATGTCAGGCTCAAGGCTTCAATCAAGTACGTGT 223
Db 984 ProValSerProProAlaProLysSerProProAlaProMetSer 1003
QY 224 CCGGATCCCGACTACTCAAGC-----AGTCCCTCCCGAGGGTTCACCT----- 271
Db 1004 ProProGluProValLysSerProProAlaProValSerProProPro 1023
QY 272 -----GGGAGCGCACCACTACGAGGACGGCGCTTCTGACCGCCACCAAGCA 325
Db 1024 ProValLysSerProProAlaProValSerProProProValLysSer 1043
QY 326 CCTCCCTGGACGGCGACTGCTGCTGTTACAAGGTGAAGATCTCGGCAACAACCTCC 385
Db 1044 Pro-----Pro-----Pro 1045
QY 386 CCGAGCGCCCGTATGACGAGNACAGCGCGCGCTGGAGCCCTCCACGAGATCGTGT 445
Db 1046 ProProAlaProValSerProProProValLysSerProProProAla--- 1064
QY 446 ACGAGGTGGACGGCGTGTGTCGCGCCAGTCCCTGATGGCCCTGAGTGGCCCGGGTC 505
Db 1065 -----ProLysSerProProPro 1071
QY 506 GCACTGACCTGACCTGACACCACTACCGCTCCA---AGAGCCCGCTCGCCCT 562
Db 1072 -----ProValLysSerProProProAlaProValSerProProPro 1088
QY 563 TGAAGATGCCCGCTTCCACTTCGAGGACCAACCGCATCGAGATCTTGGAGGAGTGA 622
Db 1089 VallysSerPro-----Pro----- 1092
QY 623 AGGCAAGTGCTACAGCAGTACGAGGCGCGCTGGGCGCTACTGCGAGCGCCCGCT 682
Db 1093 -----ProProAlaProValSerProProPro 1103
QY 683 CCA 685
Db 1104 Pro 1104
```

RESULT 8

F86387

probable Pto kinase interactor [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F86387

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chain, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F86387

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-760 <STO>

A:Cross-references: GB:AE005172; NID:g11079512; PIDN:AAG29223.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Alignment Scores:

Pred. No.: 5,89e-07 Length: 760

Score: 218.00 Matches: 75

Percent Similarity: 38.59% Conservative: 18

Best Local Similarity: 31.12% Mismatches: 85

Query Match: 16.77% Indels: 64

DB: 2 Gaps: 11

US-10-081-864-21 (1-707) x F86387 (1-760)

```
QY 20 CCGAGACCATGCTCCCTTCAGGACCA-----CCATCGAGGGCAGCCGTGA 61
Db 27 ProProGluProSerProGlyAsnAlaThrSerProThrArgLysPro--- 45
QY 62 ACGGCGACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCAGCCAGG 121
Db 46 -----ThrAsnGlyAsnProProGluThrThrAsnThrPro----- 57
QY 122 AGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181
Db 58 -----AlaGluSerProProGluThrProGluThrProLeuSerSerPro 71
QY 182 CCTCTGTCATGTACGGCTCCAGGCTTCA-----TCAAGTACGTGTGTCG 226
Db 72 ProProGluProSerProProSerProSerProSerProSerProSerProSerPro 91
QY 227 GCATCCCGACTACTTCAAGCAGTCCCTCCCGAGGGGCTTCCCTGGGAGCGCACCA 286
Db 92 ValSerProProGluProSerProProPro---ProLeuProThrGluAlaPro 110
QY 287 CTACGAGAGCGGGCTTCTGACCGCCACCGCCACAGGACACCTCCCTGGAGCGGACTGCC 346
Db 111 ProAlaAsnProValSerSerProProGluSerProProProProProProPro 128
QY 347 TGGTGTACAGGTGAAGATCTGGGGCAACACTTCCCGCGCGAGCGCCCGCTGATGCAGA 406
Db 129 -----ProThrGluAlaProProThrProThrProThrProThrSer 140
QY 407 ACAAGCGCGCGCTGGAGGAGCTCCACCGAGATCGTGTACGAGTGCAGCGGCGTGTGC 466
Db 141 ProSerProThrAsnProProProProProProProProProProProProPro 156
QY 467 GCGGCGACTCTGATGCGCTGAGTGCAGGTCGCGCGGGTGCAGCCTGACCTGCCACCTGC 526
Db 156 rleuProAlaProAspProProSerAsnProLeuProProProLysLeuValProPro 176
QY 527 ACAC-----CACCTACGCT-----CAAGAAGC-----CGGCTCGCGCTG 564
Db 176 rHisSerProProArgHisLeuProSerProProAlaSerGluLeuProProProAr 196
```


QY 565 AAGATCCCGGCTTCCACTTCGAGGACCCGCGCATCGAGATCCTCGGAGGATGGAGAG 624
 Db 196 gHisLeuProSerProPro----- 202
 QY 625 GGCAAGTGTACAAAGCAGTACGAGGCGCGTGGCGGCTACTTGCACGCGCGCCCTCC 684
 Db 203 -AlaSerGluArgProSerThrProProSerAspSerGluHisProSerProProPro 222
 QY 685 A 685
 Db 222 o 222

RESULT 9

T09854
 C:Species: Gossypium hirsutum (upland cotton)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T09854
 R:John, M.B.; Keller, G.
 Plant Physiol. 108, 669-676, 1995
 A>Title: Characterization of mRNA for a proline-rich protein of cotton fiber.
 A/Reference number: Z16888; MUID:95334493; PMID:7610164
 A/Accession: T09854
 A/Status: preliminary; translated from GE/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-214 <JOH>
 A/Cross-references: EMBL:L47308; NID:9435038; PIDN:AAA79360.1; PID:9435039
 A/Experimental source: strain Coker 312; fiber
 C/Superfamily: proline-rich protein
 C/Keywords: cell wall

Alignment Scores:
 Pred. No.: 2,31e-06 Length: 214
 Score: 209.00 Matches: 75
 Percent Similarity: 39.57% Conservative: 16
 Best Local Similarity: 32.61% Mismatches: 69
 Query Match: 16.08% Indels: 70
 DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x T09854 (1-214)

QY 32 CCTTCAGGACCATCGAGGCGCGTGAACGCCACTACTTCAAGTCAACCGCAAGG 91
 Db 26 ProSerAsnProThrSerThrPro-----AlaThrProThrProAlaSer 42
 QY 92 GCGAGGCAACCCCTCGAGGCGCACCGAGGATGAAGATCGAGTGCAGGGCGGCC 151
 Db 43 ThrProProThrThrGlnAlaPro----- 51
 QY 152 CCCTGCCCTTCGCTTCACATCTCTGTCACCTCTGCATGTACGGCTCCAGGCTTCA 211
 Db 52 -----ProThrPro-----ThrAlaThrProProProValSerThrProPro 67
 QY 212 TCAAGTACGTGTCG-----GCATCCCGACTACTTCAAGCAGTCCCTCCCGAGGCT 265
 Db 68 SerSerProProValThrAlaSerProProProValSerThrProProSerSer 87
 QY 266 TCACCTGGAGCGCACACACCTACGAGGACGGCGGCTTCTGACCGCCACGAGACA 325
 Db 88 ProProAlaThrProProPro-----AlaSerProProProAlaThrPro 103
 QY 326 CCTCCCTGGAGCGGACTGCTGTTACAGGTGAAGATCCTGGGCAACAACCTCCCG 385
 Db 104 ProPro-----AlaSerPro 108
 QY 386 CCGACGGCCCGTGTATGAGAACAGGCGCGCTGGAGCCCTCCACCGAGATCGTGT 445
 Db 109 ProProAla-----ThrProProAlaSerProProPro----- 120
 QY 446 ACAGGTGGAGCGGTGCTGGCGCGCAGTCCCTGATGGCCCTGGAGTCCCGCGGCGTC 505
 Db 121 -----AlaThrPro-----ProProAlaThrProProPro 130

QY 506 GCCACTGACCTGCCACCTGCACACACCTACCTCCGCTCCAAGAGCCGCTCCG----- 559
 Db 131 AlaThr-----ProProProAlaThrProProAlaProLeuAlaSerProProAlaThr 149
 QY 560 ---CCCTGAAGATGCCCGCTTCCACTTCGAGGACCCGCGCATCGAGATCTCGAGGAGG 616
 Db 150 ValProAlaIleSerProValGlnThrProLeuThrSerProPro----- 164
 QY 617 TGGAGAGGGCAAGTGTACAAAGCAGTACGAGGCGCGCTGGCGGCTACTTCGCGAGCGG 676
 Db 165 -----AlaProProThrGluAlaProAlaProThrLeuGlyAla---AlaThrPro 180
 QY 677 CCCCCTCAAGCTGGGCCCAACTGAAGCT 706
 Db 181 GlyProAlaGlyThrAspThrSerGlyAla 190

RESULT 10

D96711
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: D96711
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: AB6141; MUID:21016719; PMID:11130712
 A/Accession: D96711
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-708 <STO>
 A/Cross-references: GB:AE005173; NID:95734709; PIDN:AAD49974.1; GSPDB:GNO0141
 C/Genetics:
 A/Gene: F24J5.8
 A/Map position: 1

Alignment Scores:
 Pred. No.: 2,32e-06 Length: 708
 Score: 209.00 Matches: 92
 Percent Similarity: 36.75% Conservative: 19
 Best Local Similarity: 30.46% Mismatches: 106
 Query Match: 16.08% Indels: 85
 DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x D96711 (1-708)

QY 5 CCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCATCCATCGAGGCA----- 55
 Db 33 ProProValThrSerProLeuProSerAlaProProAsnArgAlaProProPro 52
 QY 56 CCGTGAACGGCCACTACTTCAAGTGCACCGGAGGCGGAGGCAACCCCTCGAGGCA 115
 Db 53 ProProProValThrSerProProProValAlaAsnGlyAlaProProProLeu 72
 QY 116 CCACGAGATGAAGATCGAGG-----TGATCGAGGCGGCCCTTCGCTTCGCT 166
 Db 73 ProLysProProGluSerSerSerProProProGlnProValIleProSerProPro 92
 QY 167 TCACATCTCTGTCACCT---CCTGCATGTACGGCTCCAGGCTTCATCAAGTACGTGT 223
 Db 93 SerThrSerProProProGlnProValIleProSerProProProSerAlaSerProPro 112
 QY 224 CCGGCA-----TCCCGGACTTACTCAAGCAGTCCCTCCCG 259
 Db 113 ProAlaLeuValProProLeuProSerSerProProProProAlaSerValProProPro 132

Db 742 Pro 742
RESULT 14
S50754
Hypothetical protein WP6 - Chlamydomonas eugametos
C:Species: Chlamydomonas eugametos
C>Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S50754
R:Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Haring,
Plant Mol. Biol. 26, 947-960, 1994
A>Title: Domain conservation in several volvoclean cell wall proteins.
A:Reference number: S50754; MUID:95093034; PMID:8000007
A:Accession: S50754
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-351 <WOF>
A:CROSS-references: EMBL:L29028; NID:g530877; PIDN:AAB53954.1; PID:g530878
Alignment Scores:
Pred. No.: 5.35e-06 Length: 351
Score: 203.50 Matches: 71
Percent Similarity: 42.86% Conservative: 22
Best Local Similarity: 32.72% Mismatches: 75
Query Match: 15.65% Indels: 49
DB: 2 Gaps: 10
US-10-081-864-21 (1-707) x S50754 (1-351)
QY 5 CCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACACCATCGAGGCGACCGTGAACG 64
Db 170 ProThrProSer---ProSerProSerProSerProSerProSerProSerProSerPro 168
QY 65 GCCACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTCGAGGCGACCCAGAGA 124
Db 189 SerProLysAlaSerProSerProSerProLysAlaSerProSerProSerProSerPro 206
QY 125 TGAAGATCGAGGTGATCGAGGCGGCGCCCTTCAGTACGCTGCTCCACATCTGTCACCT 184
Db 207 -----LysAlaSerProSerProSerProSerProSerProSerProSerProSerPro 220
QY 185 CCTCATGTACGGCTCCAGGCGCTTCATCAAGTACGTGTCGC---GCATCCCGGACTACT 241
Db 221 SerProGlnProSerProThrProSerProLysAlaSerProValAlaSerProGlnGln 240
QY 242 TCAAGCAGTCCCTCCCGAGGCTTCACCTGGGAGCGCACACCATCAGAGAGC--- 298
Db 241 SerProThrProSerProAlaGProSerProThrProSerProSerProSerProSerPro 260
QY 299 GCGGCTTCCTGACCGCCACAGGACACCTCCCTGGAGCGGACTGCTGGTGTACAAGG 358
Db 261 LysAlaSerProProProSerAlaSerPro----- 270
QY 359 TGAAGATCTGGGCAACAACCTTCCCGCGGAGCGCCCGGTGATGCGAGAACAGGCGGCG 418
Db 271 ---SerAlaSerProSerLeuSerProLysValSerPro-----SerThrProPro 286
QY 419 GCTGGGAGCCCTCCACCGAGATCTGTACGAGGTGGACGCGGTGCTGCGGCGGCGAGTCC 478
Db 287 ThrGlySerPro-----AlaAlaSerPro 294
QY 479 TGATGGCCCTGGAGTGGCGGCGGTGCGCACCTGACCTGCCACCTGCACACCATCACC 538
Db 295 -----SerGlySerProAlaSerProProGlyGlyGlyProProAlaMet 310
QY 539 GCTCCAGAGACCCGCT-----CCGCTCGAGATGCCG----- 574
Db 311 SerProArgLeuSerProThrIleProProThrArgSerProIleLeuProIleProIle 330
QY 575 GCTTCCACTTCGAGGACCGCATCGAGATCTCGGAGGAGGTGGAGAAGG 625
Db 331 ProSerProIleArgThrProSerProAlaIleGlyArgArgAsnArgArg 347

RESULT 15
T10741
extensin-like protein PRP5 - Persian tobacco
C:Species: Nicotiana glauca (Persian tobacco)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T10741
R:Schultz, C.J.; Hauser, K.; Lind, J.L.; Atkinson, A.H.; Pu, Z.Y.; Anderson, M.A.; Clark
submitted to the EMBL Data Library, May 1998
A:Reference number: Z17110
A:Accession: T10741
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-461 <SCH>
A:CROSS-references: EMBL:U88587; NID:g3116430; PID:g2653671
C:Genetics: PRP5
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein
Alignment Scores:
Pred. No.: 6.72e-06 Length: 461
Score: 202.00 Matches: 69
Percent Similarity: 35.95% Conservative: 18
Best Local Similarity: 28.51% Mismatches: 60
Query Match: 15.54% Indels: 95
DB: 2 Gaps: 8
US-10-081-864-21 (1-707) x T10741 (1-461)
QY 23 AGACCATGCCCTTCAGGACACCATCGAGGCGACCGTGAACGGCCACTTCAAGTGA 82
Db 68 LysProSerProSer---ProProAlaLysSerProProProGlnValLysSerPro 86
QY 83 CCGCAAGGGCGAGGGCAACCCCTTCGAGGCGACCCAGGAGATGAAGATCGAGGTGATCG 142
Db 87 ProProProAlaLysSerProPro----- 94
QY 143 AGGGCGGCGCCCTGCTGCTCCATCTCCATCTGCTCCACCTCTGATGATGATGATGATGAT 202
Db 95 -----ProProProAlaLysSerProProLysGlnProProProProValGlnProPro 111
QY 203 AGGCCTTCATCAAGTACGTGTCGCGCATCC----- 232
Db 112 LysGlnSerProProProSerProAlaLysGlnProProProProProProSerAlaLys 131
QY 233 ---CCGACTACTTCAAGCAGTCCCTCCCGC-----AGGCT 265
Db 132 ProProValLysProProSerProSerProAlaAlaGlnProProAlaThrGlnArgAla 151
QY 266 TCACCTGGGAGCGGACACCATCAGGAGCGGCTTCTCGACCGCCCGCCAGGAGACA 325
Db 152 ThrProProSerGlnProProProMetGlnArgAlaPro-ProProLysLeuProLeuPr 171
QY 326 CCTCCTCGAGCGGAGCTGCTGTTGATCAAGGTGAAGATCTGGGCAACAACTTCCCG 385
Db 171 oProProAlaGlnLeuProIleArgGlnProProProAlaThrGlnLeuProIle 191
QY 386 CCGAGCGGCGCGTGTATGAGAACAGGCGCGCTGGAGCCCTCCACCGAGATCGGT 445
Db 191 eArgLysPro----- 194
QY 446 ACAGGTGAGCGGTGCTGCGCGGCCAGTCCCTGATGGCCCTGGAGTGGCCCGCGGTC 505
Db 195 -----ProProProAlaGlnLeuProIleArgGln 204
QY 506 GCCACTGACCTGCGACCTGCACACCATCAGTCCAGAGCGCGCTCCCGCCCTGA 565
Db 204 mProPro-----ProProAlaThrGlnLeuProIle--ArgLysProProProPro--- 220
QY 566 AGATGCCCGGCTTCCACTTCGAGGACCCAGCATCGAGATCTCGAGGAGGTGGAGAAGG 625
Db 220 ----- 220

Search completed: July 29, 2004, 14:37:37
Job time : 22.3141 secs

Blank Sheet - US PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:24:20 ; Search time 10.2094 Seconds
(without alignments)
7211.716 Million cell updates/sec

Title: US-10-081-864-21
Perfect score: 1300
Sequence: 1 ggcacgcctccctgctgac.....ctgggcacacgaagcct 707

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2 1/USPTO spool p/US10081864/runat 29072004 150505 25463/app query.fasta_1.1742
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -IOOPCL=0 -IOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864@cgn 1.1.22 @runat 29072004 150505 25463 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	19.8	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	239.5	18.4	555	1 GPI_CHLRE	Q9fpg6 chlamydomon
3	229	17.6	620	1 EXTN_TOSAC	P13983 nicotiana t
4	217.5	16.7	238	1 GFP_AEQVI	P42212 aequorea vi
5	201.5	15.5	1033	1 IP2_STRCO	Q8cjg8 streptomyc
6	194.5	15.0	267	1 EXTN_MAIZE	P14918 zea mays (m
7	193	14.8	279	1 Y091_NPPOP	O10341 orgyia pseu
8	189.5	14.6	707	1 SPPO_HUMAN	P23246 homo sapien
9	186.5	14.3	1161	1 DAN4_YEAST	P47179 saccharomyc
10	183.5	14.1	283	1 EXTN_SORBI	P24152 sorghum bic
11	183.5	14.1	296	1 PMP3_MOUSE	P05143 mus musculu
12	183.5	14.1	1367	1 AMYH_YEAST	P08160 saccharomyc
13	182.5	14.0	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
14	182	14.0	297	1 SCS4_DROME	Q00725 drosophila
15	182	14.0	1046	1 IP2_STRAW	Q02k53 streptomyc
16	179.5	13.8	802	1 ENAH_MOUSE	Q03173 mus musculu
17	178	13.3	1733	1 VNVA_PRVKA	P33485 pseudorabie
18	176.5	13.6	1233	1 MUSA_HUMAN	P98088 homo sapien

19	175	13.5	964	1 IP2_COREF	Q8fpa7 corynebacte
20	174	13.0	1033	1 IP2_SURCO	Q8cig8 streptomyc
21	172	13.2	860	1 YHLI_EBV	P03181 epstein-bar
22	170.5	13.1	263	1 ICF3_HSV1F	P08353 herpes simp
23	170	13.1	1125	1 IP2_PROMM	Q745m4 prochloroco
24	169.5	13.0	699	1 VGLG_HSV2H	P13290 herpes simp
25	169	13.0	261	1 PRP2_MOUSE	P05142 mus musculu
26	169	13.0	426	1 EXLP_TOSAC	Q03211 nicotiana t
27	169	13.0	670	1 SYN1_MOUSE	Q88935 mus musculu
28	169	13.0	2715	1 MLL4_HUMAN	Q9unn6 homo sapien
29	168.5	13.0	1009	1 IF2_CAUCR	Q9ac25 caulobacter
30	166.5	12.8	534	1 APG_ARATH	P40602 arabidopsis
31	166.5	12.8	817	1 VRP1_YEAST	P37370 saccharomyc
32	166.5	12.8	5703	1 MUSB_HUMAN	Q9hc84 homo sapien
33	166	12.4	1046	1 IP2_STRAW	Q8k53 streptomyc
34	165.5	12.7	1185	1 DR5E_HUMAN	P54259 homo sapien
35	165	12.7	704	1 SYN1_RAT	P09951 rattus norv
36	164	12.6	437	1 YGY3_HALSQ	P21561 haloferax s
37	164	12.6	3149	1 TEGU_EBV	P03186 epstein-bar
38	163.5	12.6	1004	1 IF2_CORGL	Q8mp40 corynebacte
39	163	12.5	331	1 PRP1_HUMAN	Q04280 homo sapien
40	162.5	12.5	503	1 WAP1_HUMAN	O43516 homo sapien
41	162.5	12.5	1003	1 MED6_HUMAN	Q96dn6 homo sapien
42	162	12.5	400	1 MUAL_XENLA	P10667 xenopus lae
43	161	12.4	505	1 WASL_BOVIN	Q95107 bos taurus
44	161	12.0	641	1 EBN1_EBV	P03211 epstein-bar
45	161	12.4	917	1 SMOO_HUMAN	P53814 homo sapien

ALIGNMENTS

RESULT 1
MUC2_HUMAN STANDARD; PRT; 5179 AA.
ID MUC2_HUMAN Q02817; Q14878;
AC 002817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GS MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucus membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious

agents at mucosal surfaces.
 CC -!- SUBUNIT: Multimeric.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
 CC BRONCHUS, CERVIX AND GALL BLADDER.
 CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 VWC domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L21998; AAA95295.1; -;
 CC EMBL; M74027; AAA59875.1; -;
 CC EMBL; M94131; AAA59163.1; -;
 CC EMBL; M94132; AAA59164.1; -;
 CC PIR; A49963; A43932;
 CC Genew; HGNC: 7512; MUC2.
 CC MIM; 158370; -;
 CC GO; GO:0005803; C:secretory vesicle; TAS.
 CC InterPro; IPR006208; Cys knot.
 CC InterPro; IPR006207; Cys knot C.
 CC InterPro; IPR006209; EGF like.
 CC InterPro; IPR002919; TIL-Cysrich.
 CC InterPro; IPR001007; VWC C.
 CC InterPro; IPR001846; VWF D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; vwc; 1.
 CC Pfam; PF00094; vwd; 4.
 CC SMART; SM00214; VWC; 2.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS00022; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; VWF_1; 2.
 CC PROSITE; PS0184; VWF_2; 2.
 CC Glycoprotein; Repeat; Signal.
 CC SIGNAL 1 20
 CC CHAIN 21 5179
 CC DOMAIN 1401 1747
 CC REPEAT 1401 1416
 CC REPEAT 1417 1432
 CC REPEAT 1433 1448
 CC REPEAT 1449 1464
 CC REPEAT 1465 1471
 CC REPEAT 1472 1478
 CC REPEAT 1479 1494
 CC REPEAT 1495 1517
 CC REPEAT 1518 1533
 CC REPEAT 1534 1556
 CC REPEAT 1557 1572
 CC REPEAT 1573 1596
 CC REPEAT 1597 1612
 CC REPEAT 1613 1635
 CC REPEAT 1636 1651
 CC REPEAT 1652 1675
 CC REPEAT 1676 1693
 CC REPEAT 1694 1699
 CC REPEAT 1700 1715
 CC REPEAT 1716 1731
 CC REPEAT 1732 1747
 CC REPEAT 1748 1763
 CC DOMAIN 4815 4886
 CC VWF 1.
 CC VWF 2.
 CC 4991

FT DOMAIN 5075 5160
 FT DISULFID 5075 5122
 FT DISULFID 5089 5136
 FT DISULFID 5098 5152
 FT DISULFID 5102 5154
 FT DISULFID ? 5159
 FT CARBOHYD 163 163
 FT CARBOHYD 423 423
 FT CARBOHYD 670 670
 FT CARBOHYD 770 770
 FT CARBOHYD 894 894
 FT CARBOHYD 1139 1139
 FT CARBOHYD 1154 1154
 FT CARBOHYD 1215 1215
 FT CARBOHYD 1230 1230
 FT CARBOHYD 1246 1246
 FT CARBOHYD 1787 1787
 FT CARBOHYD 1820 1820
 FT CARBOHYD 4339 4339
 FT CARBOHYD 4351 4351
 FT CARBOHYD 4362 4362
 FT CARBOHYD 4373 4373
 FT CARBOHYD 4422 4422
 FT CARBOHYD 4438 4438
 FT CARBOHYD 4502 4502
 FT CARBOHYD 4616 4616
 FT CARBOHYD 4627 4627
 FT CARBOHYD 4752 4752
 FT CARBOHYD 4787 4787
 FT CARBOHYD 4881 4881
 FT CARBOHYD 4888 4888
 FT CARBOHYD 4955 4955
 FT CARBOHYD 4970 4970
 FT CARBOHYD 5019 5019
 FT CARBOHYD 5038 5038
 FT CARBOHYD 5069 5069
 FT CONFLICT 1351 1351
 FT CONFLICT 1412 1412
 FT CONFLICT 1449 1449
 FT CONFLICT 1504 1504
 FT CONFLICT 4192 4192
 FT CONFLICT 5179 AA; 540295 MM; 85CD7571FB9A5663 CRC64;
 FT SEQUENCE 5179 AA; 540295 MM; 85CD7571FB9A5663 CRC64;
 SQ
 Alignment Scores:
 Pred. No.: 2.17e-07 Length: 5179
 Score: 257.00 Matches: 83
 Percent Similarity: 39.93% Conservative: 24
 Best Local Similarity: 30.97% Mismatches: 93
 Query Match: 19.77% Indels: 68
 DB: 1 Gaps: 13
 US-10-081-864-21 (1-707) x MUC2_HUMAN (1-5179)
 Qy 5 CCGCCT-----CCCTGCTGACGAGACCATGCTTCCAGGACCCATCCAGG--- 52
 Db 1495 ProProThrThrThrProSerProMetThrThrProIleThrProAlaSerThr 1514
 Qy 53 -----GCACGTGAACGCCACTACT----- 73
 Db 1515 ThrThrLeuProProThrThrThrProSerProProThrThrThrThrThrProPro 1534
 Qy 74 ---TCAAGTGCACCGGAGGGGCAACCCCTCGAGGGCACCCAGGAGATGAAGA 130
 Db 1535 ProThrThrThrProSerProProThrThrThrProIleThrProThrSerThrThr 1554
 Qy 131 TCGAGGTGATCGAGGGCGCCCGCTTGCCTTCCACATCTCTGCTCCACTCTGGA 190
 Db 1555 ThrLeuProProThrThrThrProSerProProThrThrThrThrThrProProPro 1574
 Qy 191 TGTACGGCTCCAGGGCTTCATCAAGTAGTGTGCGGATCCCGACT---ACTTCAAGC 247
 Db 1575 ThrThrThrProSerProProThrThrThrThrThrProProThrThrThrThrThr 1594


```

QY 593 ACCGATCGAGATCTCTGAGGAGGCTGGAGAGGCAAGTGTCTCAAGCAGTACGAGCGC 652
Db 238 -----ProAlaProProSerProValPro 246

QY 653 CGTGGGCGGTACTGCGAGCGCGCCCTCAAGC 688
Db 247 ProSerProAlaProProSerProAlaProProSer 258

RESULT 3
EXTN_TOBAC
ID EXTN_TOBAC STANDARD; PRT; 620 AA.
AC P13983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
GN HRPNT3.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi; TISSUE=Leaf;
RX MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "Specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation.";
RL Genes Dev. 3:1639-1646 (1989).
CC -!- FUNCTION: Has a specialized structural function, possibly in
CC the mechanical penetration of the cortex and epidermis of the
CC main root.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13885; CAA32090.1; -.
DR PIR; S06733; S06733.
DR InterPro; IPR000480; Glutelin.
DR PRINTS; PR00211; GLUTELIN.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 20
FT CHAIN 21 620
FT REPEAT 70 73
FT REPEAT 148 151
FT DOMAIN 229 242
FT REPEAT 229 235
FT REPEAT 236 242
FT REPEAT 205 620
FT DOMAIN 499 600
SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Alignment Scores:
Pred. No.: 7.84e-06 Length: 620
Score: 229.00 Matches: 81
Percent Similarity: 40.23% Conservative: 22
Best Local Similarity: 31.64% Mismatches: 89
Query Match: 17.62% Indels: 64
DB: 1 Gaps: 11

US-10-081-864-21 (1-707) x EXTN_TOBAC (1-620)

```

```

QY 5 CGCCTCTCTGCTGACCGAGACCATCCCTTCAGGACCAACATCGAGG----- 52
Db 207 ProProProThHisValGlnProThrProSer---ProProSerArgGlyHisGlnPro 225

QY 53 -----GCACCTGAAAGCGGCATCTTCAAGTGCACCGCAAGG 91
Db 226 GlnProProThHisArgHisAlaPro-----ProThrHisArgHisAlaProProThr 243

QY 92 GCGAGGCAACCCCTCGAGG-----GCACCCAGGAGATGAAGATCGAGGTGA 139
Db 244 HisGlnProSerProLeuArgHisLeuProProSerProArgArgGlnProGlnProPro 263

QY 140 TCGAGGCGG-----GCCCTCGCTTCCCTTCACATCC 175
Db 264 ThrTyrSerProProProAlaTyrAlaGlnSerProGlnProSerProThrTyrSer 283

QY 176 TGTCCACCTCTGCTGATGAGGCTCCAGGCTCATCAAGTGTCCGCGCATCCCG 235
Db 284 ProProProProThrTyrSerProProProProSerProSerProSerProSerProPro 303

QY 236 ACTACTTCAAGCAGTCCCTCCCGAGGGCTTCCACCTGGGAGCGCACCACTACGAGG 295
Db 304 AlaTyrSerProSerProProProThrProThrPheSerProPro----- 321

QY 296 ACGGCGGCTTCTGACCGCCACAGGACACTCCCTGGAGCGGACTCCCTGGTGATCA 355
Db 322 ProAlaTyrSerProProProThrTyrSerProPro----- 333

QY 356 AGGTGAAGATCTCTGGGCAACAATTCCTCCCGCGAGCGCCCTGATGAGAACAGGCGC 415
Db 334 -----ProProThrTyrLeuProLeuProSerProSerPro 344

QY 416 GCGGTGGAGCGCTTCCACCGAGATGTGTACGAGGTGGAGCGGCTGTGCGGCGCAGT 475
Db 345 IleTyrSerProProPro-----ValTyrSer 354

QY 476 CCCTGATGGCTGGAGTGGCGCGCTGCACCTGACCTGCCACCTGCACACCACT 535
Db 355 ProProProPro-----ProSerTyrSerProProProProProThrTyrLeuProPro 371

QY 536 ACCGCTCCAAGAGCGCCCTCCCGCTCAAGATCCCGGCTTCCACTTCGAGGACCA 595
Db 372 ProProProSerProProProSerPheSerProProProProThrTyrGluGln 391

QY 596 GCA-----TCAGATCTGGAGAGGTGGAGAGGCAAGTGTCTACAGCAGTACG 646
Db 392 SerProProProProAlaTyrSerProProLeuProAlaProProThrTyrSerPro 411

QY 647 AGGCGCGCGTGGCGCGCTACTGCGAGCGCGCCCTCCAAAGCTGGGCC 694
Db 412 ProProPro-----ThrTyrSerProProProProThrTyrAla 424

RESULT 4
GFP_AEQVI
ID GFP_AEQVI STANDARD; PRT; 238 AA.
AC P42312; Q17104; Q27903;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";

```


FT STRAND 92 100
 FT TURN 101 102
 FT STRAND 105 115
 FT TURN 116 117
 FT STRAND 118 128
 FT TURN 132 133
 FT STRAND 135 139
 FT TURN 141 141
 FT STRAND 148 155
 FT TURN 156 159
 FT STRAND 160 171
 FT TURN 172 173
 FT STRAND 176 187
 FT TURN 199 208
 FT STRAND 211 212
 FT TURN 217 227
 SQ SEQUENCE 238 AA; 26986 MW; EA5A6P21PFB6E05 CRC64;

Alignment Scores:
 Pred. No.: 3,43e-05 Length: 238
 Score: 217.50 Matches: 59
 Percent Similarity: 47.00% Conservative: 43
 Best Local Similarity: 27.19% Mismatches: 78
 Query Match: 16.73% Indels: 38
 DB: 1 Gaps: 7

US-10-081-864-21 (1-707) x GFP_AEQVI (1-238)

QY 13 CTGTGACGAGACCATGCTCCCTTCAGGACACCATCGAGGACCGTGAACGCCACTAC 72
 Db 7 LeupheThrGlyValProileuValGluLeuAspGlyAspValAsnGlyHisLys 26
 QY 73 TTCAGTCCACCGGAGGCGGCGGACCCCTCGAGGACCGGACCGAGGATGAAGATC 132
 Db 27 PheSerValSerGlyGluGlyGlyAspAlaThrTyrGlyLysLeuThrLeuLysPhe 46
 QY 133 GAGGTGATGAGGCGGCGGCTCCCTTCGCTCCCTTCACATCCCTGCTCCATCGCATG 192
 Db 47 ---lleCysThrThrGlyLysLeuProValProThrProThrLeuValThrThrPheSer 65
 QY 193 TAGCGCTCAAGGCTTCATCAAGTACGTGTCGGCATGCCCTCCCTGACTACTCAAG 246
 Db 66 TyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAspPheLys 85
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGCGGACCCACCTACGAGGCGGCGGCTTC 306
 Db 86 SerAlaMetProGluGlyTyrValGlnGluArgThrLeuPhePheLysAspAspGlyAsn 105
 QY 307 CTGACCGCCCGGACGACACCTCCCTGAGCGGAGTCCCTGTTCAAGGTGAAGATC 366
 Db 106 TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 125
 QY 367 CTGGGACACACTCCCGCGGCGGCGGCTGATGAGACAG-- 411
 Db 126 LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyr 145
 QY 412 -----GCCGCGCGTGGGAGCGCTCCACCGAGATCGTGATC 447
 Db 146 AsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 165
 QY 448 GAGGTG-----GAGGCGGTGTCGCGCG-- 470
 Db 166 LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn 185
 QY 471 ---CCAGTCCCTGATGGCGCTGGAGTCCCGCGGCGGCTCCACCTGACCTGCCACCTGCA 527
 Db 186 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 198
 QY 528 CACACCTACCGCTCCCAAGAGCGCGCTCCGCG---CCTGAAGATGCCCGG 575
 Db 199 HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg 215

RESULT 5

IF2_STRCO
 ID IF2_STRCO STANDARD; PRT; 1033 AA.
 AC Q8CJQ8;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Translation initiation factor IF-2.
 GN INFB OR SC05706 OR SC9F2.10C OR SC5H4.30.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2) / M145;
 RX MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)"; (2002).
 RL Nature 417:141-147 (2002).
 CC -!- FUNCTION: One of the essential components for the initiation of
 CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
 CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
 CC Also involved in the hydrolysis of GTP during the formation of the
 CC 70S ribosomal complex (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the IF-2 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL939124; CAD55362.1; --
 DR PIR; T35989; T35989.
 DR HAMAP; MF_00100; -- 1.
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR001851; Bac_innem_transp.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR000795; EF_GTPbind.
 DR InterPro; IPR000178; IF2.
 DR InterPro; IPR006847; IF2_N.
 DR InterPro; IPR001806; Ras_transfmg.
 DR InterPro; IPR005225; Small_GTP.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR Pfam; PF04760; IF2_N; 2.
 DR PRINTS; PR00929; ATHOOK.
 DR PRINTS; PR00315; ELONGATNFACT.
 DR PRINTS; PR00449; RASTRNSFRMG.
 DR Prodom; PD186100; IF2; 1.
 DR TIGRFAMs; TIGR00487; IF-2; 1.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 DR PROSITE; PS01176; IF2; FALSE NEG.
 KW Initiation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT DOMAIN 529 681 G-DOMAIN
 FT NP_BIND 535 542 GTP (BY SIMILARITY).
 FT NP_BIND 585 589 GTP (BY SIMILARITY).
 FT NP_BIND 639 642 GTP (BY SIMILARITY).
 SQ SEQUENCE 1033 AA; 105657 MW; 47D581PFB0072A045 CRC64;

binding protein-associated splicing factor) (PTB-associated splicing factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
 GN SFPQ OR PSF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal brain;
 RX MDLINE=93194059; PubMed=8449401;
 RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
 RT "Cloning and characterization of PSF, a novel pre-mRNA splicing factor.";
 RT factor.";
 RL Genes Dev. 7:393-406(1993).
 RN [2]
 RP SEQUENCE OF 312-707 FROM N.A.
 RC TISSUE=Fetal skeletal muscle;
 RX MDLINE=90091812; PubMed=2480877;
 RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
 RT "Cloning and characterization of a myoblast cell surface antigen defined by 24.1D5 monoclonal antibody.";
 RL Development 105:723-731(1989).
 RN [3]
 RP SEQUENCE OF 48-68 AND 213-246.
 RX MDLINE=93176127; PubMed=8439294;
 RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
 RT "Purification and characterization of a DNA-binding heterodimer of 52 and 100 kDa from HeLa cells.";
 RL Biochem. J. 290:267-272(1993).
 CC -1- FUNCTION: Essential pre-mRNA splicing factor required early in spliceosome formation. Binds to the mammalian polypyrimidine tracts. Forms a complex with the polypyrimidine tract-binding protein (PTB). Seems to also bind DNA.
 CC -1- SUBUNIT: Heterotetramer of two 52 kDa and two 100 kDa subunits.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=P23246-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P23246-2; Sequence=VSP_005855;
 CC -1- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
 CC -1- CAUTION: Was originally (Ref.2) thought to be myoblast cell surface antigen 24.1D5 and a possible membrane-bound protein ectokinase.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X70944; CAA50283.1; -;
 CC EMBL; X16850; CAA34747.1; -;
 CC PIR; A46302; A46302.
 CC PIR; S29770; S29770.
 CC HSP; P11940; IGVJ.
 CC SWISS-2DPAGE; P23246; HUMAN.
 CC Genew; HGNC:10774; SFPQ.
 CC GK; P23246; -;
 CC MIM; 605199; -;
 CC GO; GO:0008248; P:pre-mRNA splicing factor activity; TAS.
 CC GO; GO:0006397; P:mRNA processing; TAS.
 CC GO; GO:0006371; P:mRNA splicing; TAS.
 CC InterPro; IPR000504; RNA_rec_mot.
 CC Pfam; PF00076; rrm; 2.
 CC SMART; SM00360; RRM; 2.
 CC PROSITE; PS01012; RRM; 2.
 CC PROSITE; PS00030; RRM_RNP_1; 1.

KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
 KW Alternative splicing.
 FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
 FT 371 452 RNA-BINDING (RRM) 2.
 FT 9 27 3 X 3 AA REPEATS OF R-G-G.
 FT REPEAT 9 11 1.
 FT REPEAT 19 21 2.
 FT REPEAT 25 27 3.
 FT 266 663 GLN/GLU/PRO-RICH.
 FT 10 15 POLY-GLY.
 FT 20 27 POLY-GLY.
 FT 56 65 POLY-PRO.
 FT 67 71 POLY-GLN.
 FT 95 98 POLY-GLN.
 FT 99 103 POLY-PRO.
 FT 184 188 POLY-PRO.
 FT 571 574 POLY-ARG.
 FT 613 616 POLY-GLY.
 FT 635 641 POLY-GLY.
 FT 663 707 VARSPLIC
 FT
 FT RTERFGGGAGPVGSGQPRGMGPCTPAGYGRGREYEGPNK
 FT XPRF -> VRMDVIG (in isoform Short).
 FT /FTid=VSP_005855.
 FT G -> R (in REF. 3).
 SQ CONFLICT 243 243
 SQ SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;
 Alignment Scores:
 Pred. No.: 0.00117 Length: 707
 Score: 189.50 Matches: 75
 Percent Similarity: 37.21% Conservative: 21
 Best Local Similarity: 29.07% Mismatches: 91
 Query Match: 14.58% Indels: 71
 DB: 1 Gaps: 10
 US-10-081-864-21 (1-707) x SFPQ_HUMAN (1-707)
 QY 5 CGCGCTCTGCTGACCGAGACCATGCCCTTCAGGACACCATCGAGGCGACCGTGAACG 64
 Db 72 ProProPro-----GlnGlnProProGlnGlnProPro-----HisGlnProPr 88
 QY 65 GCCTACTTCAAGTGACCGGCAAGGCGAGGCGCAACCCCTCGAGGCGACCCAGGAGA 124
 Db 88 oProHisProGlnProHisGlnGln-----GlnGlnProPro-ProProGlnAsps 106
 QY 125 TGAAGATCGAGTGATGACGAGGCGCGCCCTCGCTTCACATCTCCATCTGCTCCACT 184
 Db 106 exSerLysProValValAlaGlnGly-ProGlyProAlaProGlyValGlySerAlaPro 125
 QY 185 CTGCTGATGACGGCTCCAGGCGCTTCATCAAGTAGCTGTCGCGCATCCCGACTACTTCA 244
 Db 126 ProAla-----SerSerSerAlaProProAlaProProProThrSer 139
 QY 245 AGCAGTCTCCCTCCGAGGCGCTTCCTGGAGCGGCGACCACTTACGAGGCGCGCT 304
 Db 140 GlyAlaPro---ProGlySerGlyProGlyProThrProThrProProAlaValThr 158
 QY 305 TCTGACGCGCCACGAGGACACTCCCTGGAGCGGCGACTGCTGCTGTACAGGTGAAGA 364
 Db 159 SerAlaProProGlyAlaProProPro-----
 QY 365 TCTGGGCAACAACACTCCCGCGCGCGCGCGCGTATGCAGAACAGCGCGCGCTGGG 424
 Db 168 -----ThrProSerSerGlyValProThrProGlnAlaGly 182
 QY 425 AGCCTCCACCGAGATCTGTGA-----
 Db 183 GlyProProPro-ProProAlaAlaValProGlyProGlyProGlyProLysGlnGlyPr 202
 QY 447 -----CGAGGTGGACGGCGTCTCGCGCGCGCGCTCCCT---GATGGCCC 487
 Db 202 oGlyProGlyProLysGlyGlyLysSerProGlyGlyProLysProGlyGlyGlyPr 222
 QY 488 TGGAGTGGCGCGCGCTGCCACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 524
 |||||

Db 222 OGlyLeuSerThrProGlyGlyHisProLysProHisArgGlyGlyGluProAr 242
 QY 525 -----GCACACCACTACCGCTCCAGAGAGCCCGCTCCGCTGAGATGCCG 574
 Db 242 GOGlyArgGlnHisProProTyHisGlnGlnHis ----- 255
 QY 575 GCTTCCACTTCGAGGACCAACCGCATCGAGATCCTGGAGAGGTGAGAGAGGCAAGTGCT 634
 Db 256 -----HisGlnGlyProProGlyGlyProGlyGlyArgSerGluGluLysIleSe 273
 QY 635 ACAACGACATAGAGCGCCCGCTGGCGCTACTCGAGCGCGCC 680
 Db 273 rAspSerGluGlyPheLysAlaAsnLeuSerLeuLeuArgPro 288

RESULT 9

DN4_YEAST
 ID DN4_YEAST STANDARD; PRT; 1161 AA.
 AC P47179;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cell wall protein DN4 precursor.
 GN DN4 OR YR151C OR J2223.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Scarcez T.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REGULATION.
 RX MEDLINE=21113168; PubMed=11160904;
 RA Cohen B.D., Sertliff O., Abramova N.E., Davies K.J., Lowry C.V.;
 RT "Induction and repression of DAN1 and the family of anaerobic
 RT mannoprotein genes in saccharomyces cerevisiae occurs through a
 RT complex array of regulatory sites."
 RL Nucleic Acids Res. 29:799-808(2001).
 CC -!- FUNCTION: Component of the cell wall (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: Extensively O-glycosylated (Potential).
 CC -!- SIMILARITY: Belongs to the SRP1 / TIP1 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z49651; CAA89684.1; -;
 DR PIR; S57180; S57180.
 DR GenOnline; 141984; -;
 DR SGD; S0003912; DN4.
 DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
 DR InterPro; IPR000992; SRP1_TIP1.
 DR Pfam; PF00660; SRP1_TIP1; 1.
 DR PROSITE; PS00724; SRP1_TIP1; 1.
 KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal; Lipoprotein.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1146 CELL WALL PROTEIN DN4.
 FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 1146 1146 GPI-anchor amidated asparagine
 FT (Potential).
 SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D69FOCA58 CRC64;
 Alignment Scores:
 Pred. No.: 0.0017 Length: 1161
 Score: 186.50 Matches: 61
 Percent Similarity: 42.27% Conservative: 32

Best Local Similarity: 27.73% Mismatches: 92
 Query Match: 14.35% Indels: 35
 DB: 1 Gaps: 5
 US-10-081-864-21 (1-707) x DN4_YEAST (1-1161)
 QY 47 TCGAGGGCACCCTGTAACCGCCACTATTCAAGTCACCGGAGGCGGAGGCAACCC 106
 Db 150 SerThrThrProThrThrSerThrThrSerThrThrSerThrThrSerThr 168
 QY 107 TCGAGGGCACCAGAGATGAGATCGAGGGCGGCGGCGGCGGCGGCGGCT 166
 Db 169 -----ThrThrProThrThrSerThr 175
 QY 167 TCGACATCCTGTCACCTCCTGCTGCTGCTCCAGGCTTCAAGTACGTCGTCG 226
 Db 176 ThrSerThrThrProThrThrSerThrThrSerThrThrSerThrThrSer 195
 QY 227 GCATCCCGACTACTTCA---AGCAGTCCCTCCCGAGGCTTACCTGGGAGCGCACA 283
 Db 196 ThrThrProThrThrSerThrThrSerThrThrSerThrThrSerThrThr 215
 QY 284 CCACCTAGGAGGAGCGCGCTTCTGACCGCCACCGACACACCTCCCTGGACGCGACT 343
 Db 216 ProThrThrSerThrThrSerThrThrProThrThrSerThrThrSerThr 235
 QY 344 GCTGCTGTACAAGGTGAAGATCTCTGGGCAACAACCTCCCGCGGACGCGCGCTGATGC 403
 Db 236 ThrSerThrThrSerGlnThrSerThrLysSerThrThrProThrThrSer 254
 QY 404 AGAACAAGCGCGCGCTGGGAGCCCTCCACGAGATCGTGTACGAGTGGACGCGTCG 463
 Db 255 SerThrThrProThrThrSerThrThrProThrThrSerThrThrSer 270
 QY 464 TCGCGGCGCAGTCCCTGATGCGCTGGAGTGCCTCCGCGGCTGCGCACCTGACCTGCCACC 523
 Db 271 -----ThrAlaProThrThrSerThrThrSerThrThr 281
 QY 524 TCGACACACCTACCGCTCCAGAGCGCGCTCCGCGCTGAGATGCGCGCTTCCACT 583
 Db 282 SerThrThrSerThrIleSerThrAlaPro-----ThrThrSerThr 295
 QY 584 TCGAGGACCAACCGCATCGAGATCCTGGAGAGGTGGAGAGGCAAGTGTACAGCAGT 643
 Db 296 ThrSerThrThrPheSerThrSerSerAlaSerAlaSerSerValIleSerThrAla 315
 QY 644 ACAGGCGCGCGCTGGGCGCTGCTGCGACGCGCGCGCTCCAGTGGGCGGCAACTGAA 703
 Db 316 ThrThrSerThrThrPheAlaSerLeuThrThrProAlaThrSerThrAlaSerThrAsp 335

RESULT 10

EXTN_SORBI
 ID EXTN_SORBI STANDARD; PRT; 283 AA.
 AC P24152;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Extensin precursor (Proline-rich glycoprotein).
 GN HRGP.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=91370882; PubMed=1893107;
 RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;
 RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum
 RT vulgare."
 RL Plant Mol. Biol. 16:365-367(1991).
 CC -!- FUNCTION: Structural component in primary cell wall.

```
Db      196 -----ProValTyThrProSerProLysProPolysProSer 208  
Qy      665 -----ACTGCGAGCCGCCCTTCCAAGC 688  
Db      209 ProProThrTythrProThrProLysProProAla 220  
  
RESULT 11  
PMP3 MOUSE STANDARD; PRT; 296 AA.  
AC P05143;  
DT 13-AUG-1987 (Rel. 05, Created)  
DI 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 10-OCT-1983 (Rel. 42, Last annotation update)  
DE Feline-rich protein MF-3 (fragment).  
GN PRH1 OR PRP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86059475; PubMed=2999141;  
RA Ann D.K., Carlson D.M.;  
RT "The structure and organization of a proline-rich protein gene of a mouse sturgeon family."  
RL J. Biol. Chem. 260:15863-15872(1985).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).  
DR EMBL; M12100; AAA40005.1; --  
DR MGD; MGJ:97773; Ptbl.  
FT NON TER  
SQ SEQUENCE 296 AA; 29521 MW; 7F146824EBAF3269 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.00255 Length: 296  
Score: 183.50 Matches: 72  
Percent Similarity: 35.86% Conservative: 13  
Best Local Similarity: 30.38% Mismatches: 93  
Query Match: 14.12 Indels: 60  
DB: 1 Gaps: 12  
  
US-10-081-864-21 (1-707) x PMP3_MOUSE (1-296)  
Qy 23 AGACCATGCCTTTCAAGAACCCACCATCACGGCCACGTGAACGGCCACTATTCAA GTGCA 82  
Db 12 ArgProProSerArglyserGlnProAargProProVaIasnGlySerGlnGInglyPro 31  
Qy 83 CCGGCAAGGCGGAGGG-----CAACCCCCTCAGGGGCCCCAA ----GGAGA 124  
Db 32 ProProProgily-GlyProGlnProAargProProGInglyProProProgilygPr S1  
Qy 125 TGAAGATCGAGGTGATCGAGGGGGGCCCCCTGC-----CCTTCGCCTCCAC 171  
Db 51 oGlnProAargProProGInglyProProProProgilyGlyProGlnProAargProGl 71  
Qy 172 ATCCTGTCCACCTCCCTGCATGTAGCGCTCAAGGCCTTCATCAAGTCGTGTCCGCGATC 231  
Db 71 nGlyProProProProgilycylProGlnProAargPro-----ProGlnGI 86  
Qy 232 CCCGACTACTTCAAGCAGTGCTCCCGGGGGCTTCACTGGGAGCGCACCCACCTAC 291  
Db 86 yProProProProProgilycylProGlnProArg-----ProProGlnGlyProProProp 104  
Qy 292 GAGGAGCGGGCGV'TCCTGACCGGCCACAGGACACCTCCCTGGACGGCGACTGCTGGTG 351
```


Search completed: July 29, 2004, 14:33:32
Job time : 18.2094 secs

Blank Sheet U.S. PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:27:25 ; Search time 42.1137 Seconds

(without alignments)
10593.763 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300
Sequence: 1 gggatccgctccctgctgac.....ctggggccacactgaagctt 707

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2.1/USPTO.spool_p/US10081864/runat.29072004.150506.25475/app.query.fasta_1.1742
-DB=SPTREMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 -CGEN_1_122 -runat.29072004.150506.25475 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
1	1219	93.8	232	5	Q9GZ28	Q9GZ28 anemonia su

2	1186	91.2	232	5	Q9GP15	Q9GP15 anemonia su
3	781.5	60.1	227	5	Q9SW86	Q9SW86 condylactis
4	779.5	60.0	227	5	Q9SW11	Q9SW11 condylactis
5	763.5	58.7	227	5	Q9SW85	Q9SW85 radianthus
6	747.5	57.5	227	5	Q8MU45	Q8MU45 condylactis
7	739.5	56.9	231	5	Q8ISF8	Q8ISF8 parasicyoni
8	637	49.0	228	5	Q9GP16	Q9GP16 anemonia su
9	561.5	43.2	225	5	Q7ZOW9	Q7ZOW9 montastraea
10	560.5	43.1	229	5	Q8TSE7	Q8TSE7 condylactis
11	559	43.0	224	5	Q8MU48	Q8MU48 montastraea
12	559	43.0	225	5	Q7ZOW5	Q7ZOW5 montastraea
13	558.5	43.0	227	5	Q9SVT0	Q9SVT0 montastraea
14	553.5	42.6	225	5	Q8T5F1	Q8T5F1 montastraea
15	553.5	42.6	236	5	Q8T6U0	Q8T6U0 dendronephr
16	553	42.5	225	5	Q9UA7	Q9UA7 montastraea
17	551.5	42.4	221	5	Q9SP04	Q9SP04 gonopora t
18	551.5	42.4	227	5	Q962P9	Q962P9 montastraea
19	551.5	42.4	227	5	Q7ZOW8	Q7ZOW8 montastraea
20	550.5	42.3	225	5	Q7ZOW4	Q7ZOW4 montastraea
21	548.5	42.2	225	5	Q963F5	Q963F5 montastraea
22	547.5	42.1	227	5	Q7ZOW6	Q7ZOW6 montastraea
23	545.5	42.0	225	5	Q9U6V8	Q9U6V8 discosoma s
24	543	41.8	225	5	Q8T8T9	Q8T8T9 radianthus
25	541.5	41.7	234	5	Q8T5F2	Q8T5F2 montastraea
26	541.5	41.7	234	5	Q7ZOW7	Q7ZOW7 montastraea
27	541	41.6	230	5	Q9GTJ7	Q9GTJ7 discosoma s
28	536	41.2	231	5	Q8T5E8	Q8T5E8 ricordea fl
29	536	41.2	232	5	Q9U6V7	Q9U6V7 discosoma s
30	535	41.2	228	5	Q86LV4	Q86LV4 radianthus
31	529.5	40.7	266	5	Q9U6Y3	Q9U6Y3 clavularia
32	529.5	40.7	266	5	Q8T6T8	Q8T6T8 discosoma s
33	528	40.6	231	5	Q8MU46	Q8MU46 ricordea fl
34	523	40.2	231	5	Q8I6J8	Q8I6J8 trachyphyl
35	519	39.9	225	5	Q8I6J8	Q8I6J8 ricordea fl
36	513.5	39.5	231	5	Q8T5E9	Q8T5E9 ricordea fl
37	470	36.2	225	5	Q8T5F0	Q8T5F0 scolymlia cu
38	469.5	36.1	235	5	Q9U6V6	Q9U6V6 anemonia ma
39	464	35.7	234	5	Q8T5F3	Q8T5F3 scolymlia cu
40	460	35.4	259	5	Q8MMA2	Q8MMA2 agaricia fr
41	453.5	34.9	238	5	Q9BLZ0	Q9BLZ0 ptilosarcus
42	445.5	34.3	238	5	Q9BLV9	Q9BLV9 renilla mue
43	434	33.4	239	5	Q8MMA1	Q8MMA1 agaricia ag
44	423.5	32.6	214	5	Q86LV7	Q86LV7 meandrina m
45	422.5	32.5	214	5	Q86LV8	Q86LV8 meandrina m

ALIGNMENTS

Handwritten signature

RESULT 1
Q9GZ28 PRELIMINARY; PRT; 232 AA.

AC Q9GZ28; (1)
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE GFP-like chromoprotein FP595.
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actiniidae; Anemonia.
OX NCBI_TaxID=6108;
RN [1]
RP SEQUENCE FROM N.A. PubMed=10852900;
RX MEDLINE=24408921; PubMed=10852900;
RA Lukyanov K.A., Fradkov A.F., Gurskaya N.G., Matz M.V., Labas Y.A.,
RA Savitsky A.P., Markelov M.L., Zaraisky A.G., Zhao X., Fang Y., Tan W.,
RA Lukyanov S.A.;
RT "Natural Animal Coloration Can Be Determined by a Nonfluorescent Green
RT Fluorescent Protein Homolog."
RL J. Biol. Chem. 275:25879-25882 (2000).
DR EMBL; AF246709; AAG02385.1; ...
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFLUORESCENT.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 232 AA; 25918 MW; CDFE982006F4975E CRC64;

Alignment Scores:
 Pred. No.: 3,41e-73 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservative: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x O9GZ28 (1-232)

QY 7 GCCTCCCTGTCGACGAGACCATGCTTCAGGACCACTCGAGGCGACCGTCAAGCGC 66
 DB 2 AlaserPheLeuLysLysThrMetProPheLysThrThrLeuGluGlyThrValAsnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATG 126
 DB 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41

QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTCCCTTCACATCTCCATCTGTCACCTCC 186
 DB 42 LysileGluValIleGluGlyGlyProLeuProPheAlaPheHisLeuSerThrSer 61

QY 187 TGCATGTACGCTCCAAAGCGCTTCATCAAGTACGTGTCGGCATCCCGGACTACTTCAAG 246
 DB 62 CysMetTyrGlySerLysThrPheLysTyrValSerGlyLeuProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGCGCTCACTCGGAGCGGACACACCTACGAGGAGCGCGCTTC 306
 DB 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrThrThrThrThrThrThrThr 101

QY 307 CTGACCGCCACCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366
 DB 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121

QY 367 CTGGGCAACACTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
 DB 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGluArgTrpGlu 141

QY 427 CCCTCCACCGAGATCGGTGTACGAGGTGACGGCGGTGTGTGTGTGTGTGTGTGTGTGT 486
 DB 142 ProAlaThrGluLeuLeuTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161

QY 487 CTGAGTGCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 546
 DB 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisSerThrTyrArgSerLys 181

QY 547 AAGCCCGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606
 DB 547 AAGCCCGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606

QY 607 CTGAGGAGGTGGAGAGGCGCAAGTGTACAACAGTACGAGCGCGCGTGGCGCGCTAC 666
 DB 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221

QY 667 TGGACCGCGCGCTCCCAAGCTGGGCGCAAC 699
 DB 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 2
 O9GP15
 ID O9GP15 PRELIMINARY; PRT; 232 AA.
 AC O9GP15
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Nonfluorescent red protein asCp562.
 OS Anemonia sulcata (Snake-locks sea anemone).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;

OC Nymnathae; Actiniidae; Anemonia.
 OX NCBI_TaxID=6108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20570437; PubMed=11121018;
 RA Wiedenmann J., Elke C., Spindler K.D., Funke W.;
 RT "Cracks in the beta -can: fluorescent proteins from Anemonia sulcata
 (Anthozoa, Actinaria).";
 RL Proc. Natl. Acad. Sci. U.S.A. 97;14091-14096 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Wiedenmann J., Elke C., Spindler K.-D., Funke W., Sundin B., Jach G.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF322222; AAG41206.2; -
 DR HSSP; P42212; 1GFL.
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP-like.
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFLUORESCENT.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 232 AA; 25956 MW; D7B586FE548097F3 CRC64;

Alignment Scores:
 Pred. No.: 5.37e-71 Length: 232
 Score: 1186.00 Matches: 215
 Percent Similarity: 96.54% Conservative: 8
 Best Local Similarity: 93.07% Mismatches: 8
 Query Match: 91.23% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x O9GP15 (1-232)

QY 7 GCCTCCCTGTCGACGAGACCATGCTTCAGGACCACTCGAGGCGACCGTCAAGCGC 66
 DB 2 AlaserPheLeuLysLysThrMetProPheLysThrThrLeuGluGlyThrValAsnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATG 126
 DB 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41

QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTCCCTTCACATCTCCATCTGTCACCTCC 186
 DB 42 LysileGluValIleGluGlyGlyProLeuProPheAlaPheHisLeuSerThrSer 61

QY 187 TGCATGTACGCTCCAAAGCGCTTCATCAAGTACGTGTCGGCATCCCGGACTACTTCAAG 246
 DB 62 CysMetTyrGlySerLysThrPheLysTyrValSerGlyLeuProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGCGCTCACTCGGAGCGGACACACCTACGAGGAGCGCGCTTC 306
 DB 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrThrThrThrThrThrThrThr 101

QY 307 CTGACCGCCACCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366
 DB 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121

QY 367 CTGGGCAACACTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
 DB 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGluArgTrpGlu 141

QY 427 CCCTCCACCGAGATCGGTGTACGAGGTGACGGCGGTGTGTGTGTGTGTGTGTGTGT 486
 DB 142 ProAlaThrGluLeuLeuTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161

QY 487 CTGAGTGCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 546
 DB 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisSerThrTyrArgSerLys 181

QY 547 AAGCCCGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606
 DB 547 AAGCCCGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606

QY 607 CTGAGGAGGTGGAGAGGCGCAAGTGTACAACAGTACGAGCGCGGTGGCGCGCTAC 666
 DB 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221

QY 667 TGGACCGCGCGCTCCCAAGCTGGGCGCAAC 699
 DB 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

```

QY 607 CTGAGGAGGTGGAGAGGGCAAGTGTCTACAGCAGTACGAGCGCGCGTGGCGCGCTTAC 666
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
202 MetGlulValGlulysGlyCysTyrGlnTyrGluAlaValAlaArgTyr 221
QY 667 TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
222 CysAspAlaAlaProSerLysLeuGlyHisHis 232

RESULT 3
Q95W86 PRELIMINARY; PRT; 227 AA.
AC Q95W86;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Condylactis gigantea (Giant anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actiniidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Fradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins (1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF363775; AAL27537.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25416 MW; BCFAS44CBCE1B3F7 CRC64;

Alignment Scores:
Pred. No.: 4 53e-44 Length: 227
Score: 78.50 Matches: 148
Percent Similarity: 78.41% Conservative: 30
Best Local Similarity: 65.20% Mismatches: 46
Query Match: 60.12% Indels: 3
DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x Q95W86 (1-227)
QY 7 GCCTCCCTGTCAGCAGACCATGCCCTTCAGGACACCATCAGGCGCCGCTGAGCGC 66
Db 2 AlaGlyLeuLeuLysGluSerMetArgIleLysIleTyrMetGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTGCACCGGAGGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db 22 TyrHisPheLysCysGluGlyAspGlyAsnProPheAlaPheAspIleLeuSerProCys 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 42 ArgIleArgValThrGluGlyAlaProLeuProPheAlaPheAspIleLeuSerProCys 61
QY 187 TGCATGTACGGTCCCAAGGCGCTTCATAGTACGTGTCCGGCATCCCCGACTACTTCAAG 246

```

```

Db 122 LeuGlyThrAsnPheProAlaAspGlyProValMetLysLysIleSerGlyGlyTyrGlu 141
QY 427 CCTTCCACCGAGATCGTGTACGAGGTGACGGGTGCTGCGCGCGCGCGCGCGCGCGCG 486
Db 142 ProCysThrGluIleValTyrGlnAspAsnGlyValLeuArgGlyArgAsnValMetAla 161
QY 487 CTGGAGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
Db 162 LeuLysValSerGlyArgProProLeuIleCysHisLeuHisSerThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
Db 182 Lys--AlaCysAlaLeuThrMetProGlyPheHisPheAlaAspLeuArgIleGln-- 199
QY 607 CTGAGGAGGTGGAGAGGGCAAGTGTCTACAGCAGTACGAGCGCGCGCGCGCGCGCT 666
Db 200 MetProLysLysLysLysAspGlyTyrPheLeuLysLysLysLysLysLysLysLys 219
QY 667 TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
Db 220 SerAsp--ValProGluLys 225

RESULT 4
Q95W11 PRELIMINARY; PRT; 227 AA.
AC Q95W11;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Condylactis passiflora.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actiniidae; Condylactis.
OX NCBI_TaxID=175772;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Fradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins (1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF363155; AAL27541.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25446 MW; E51CC017108593E3 CRC64;

Alignment Scores:
Pred. No.: 6.16e-44 Length: 227
Score: 779.50 Matches: 149
Percent Similarity: 78.41% Conservative: 29
Best Local Similarity: 65.64% Mismatches: 46
Query Match: 59.96% Indels: 3
DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x Q95W11 (1-227)
QY 7 GCCTCCCTGTCAGCAGACCATGCCCTTCAGGACACCATCAGGCGCCGCTGAGCGC 66
Db 2 AlaGlyLeuLeuLysGluSerMetArgIleLysIleTyrMetGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTGCACCGGAGGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db 22 TyrHisPheLysCysGluGlyAspGlyAsnProTyrGluGlyThrGlnAsnMet 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 42 ArgIleArgValThrGluGlyAlaProLeuProPheAlaPheAspIleLeuSerProCys 61
QY 187 TGCATGTACGGTCCCAAGGCGCTTCATAGTACGTGTCCGGCATCCCCGACTACTTCAAG 246

```

```

Db      62 CysAlaTyrGlySerLysThrPheIleLysHisThrSerGlyIleProAspTyrPheLys 81
QY      247 CAGTCCCTCCCGAGGGGTTTCACTGGAGGCGACACCACTACGAGGACGGGGCTTC 306
Db      82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrIleTyrGluAspGlyGlyVal 101
QY      307 CTGACCGCCACCAAGGACACCTCCCTGACGCGGACTGCTGGTGTACAAAGTGAAGTC 366
Db      102 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnCysLeuAsnTyrLysValLysVal 121
QY      367 CTGGCAACAACCTTCCCGCGAGCGCCCGCTGATGAGAACAAAGCGCGCGCTGGGAG 426
Db      122 LeuGlyThrAsnProAlaAspGlyProValMetLysAsnIleSerGlyGlyTyrGlu 141
QY      427 CCCTCCACCGAGATCGTGTAGAGGTGACGGCGTGTGGCGGCGAGTCCCTGATGGCC 486
Db      142 ProCysThrGluIleValTyrGlnAspAsnGlyValLeuArgGlyArgAsnValMetAla 161
QY      487 CTGAGTGCCTCCCGCGGTGCGCCACTGACCTGCGCACCTGCGACACCACTACCGCTCCAAG 546
Db      162 LeuLysValSerGlyArgProProLeuIleCysHisLeuHisSerThrTyrArgSerLys 181
QY      547 AAGCCCGCTCCGCGCTGAGATGCGCGCTTCCACTTCGAGGACCAACCGATCGAGATC 606
Db      182 Lys---AlaCysAlaLeuThrMetProGlyPheHisPheAlaAspLeuArgIleGln--- 199
QY      607 CTGAGGAGGTGGAGAGGCAAGTGTACAAAGCAGTACGAGCGCGCGCTGGCGCGTAC 666
Db      200 MetProLysLysLysLysAspGlyTyrPheGluLeuTyrGluAlaSerValAlaArgTyr 219
QY      667 TGCAGCGCGCGCGCTCCAAG 687
Db      220 SerAsp---LeuProGluLys 225

RESULT 5
Q95W85
ID Q95W85 PRELIMINARY; PRT; 227 AA.
AC Q95W85;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Radianthus macrodactylus (Sea anemone) (Heteractis macrodactylus).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Stichodactylidae; Heteractis.
OX NCBI_TaxID=175771;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11692051;
RA Gurskaya N.G.; Fradkov A.F.; Tersikh A.; Matz M.V.; Labas Y.A.;
RA Martynov V.I.; Yanushevich Y.G.; Lukyanov K.A.; Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF363776; AAL27538.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25637 MW; CB40899B95E7EC64 CRC64;

Alignment Scores:
Pred. No.: 7.16e-43 Length: 227
Score: 763.50 Matches: 141
Percent Similarity: 77.09% Conservative: 34
Best Local Similarity: 62.11% Mismatches: 49
Query Match: 58.73% Indels: 3
DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x Q95W85 (1-227)

```

```

QY      7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGC 66
Db      2 AlaGlyLeuLeuLysGluSerMetArgLysMetTyrMetGluGlyThrValAsnGly 21
QY      67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db      22 HisTyrPheLysCysGluGlyGluGlyAspGlyAsnProPheThrGlyThrGlnSerMet 41
QY      127 AAGATCGAGGTATCGAGGGGGGGCCCTGCCCTTGGCTTCCACATCTCTCCACCTCC 186
Db      42 ArgIleHisValThrGluGlyAlaProLeuProPheAlaPheAspIleLeuAlaProCys 61
QY      187 TGCATGTACGGTCCCAAGGCGCTTCATCAAGTACGTGTCGGCATCCCGCATCTTCAAG 246
Db      62 CysGluTyrGlySerArgThrPheValHisThrAlaGluIleProAspPhePheLys 81
QY      247 CAGTCCCTCCCGAGGGCTTCACTGGGAGGCGCACCACTACGAGGACGGGGCTTC 306
Db      82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyIle 101
QY      307 CTGACCGCCACCAAGGACACCTCCCTGACGCGGACTGCTGGTGTACAAAGTGAAGTC 366
Db      102 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnCysLeuIleTyrLysValLysVal 121
QY      367 CTGGCAACAACCTTCCCGCGAGCGCCCGCTGATGAGAACAAAGCGCGCGCTGGGAG 426
Db      122 LeuGlyThrAsnProAlaAspGlyProValMetLysAsnLysSerGlyGlyTyrGlu 141
QY      427 CCCTCCACCGAGATCGTGTAGAGGTGACGGCGTGTGGCGGCGAGTCCCTGATGGCC 486
Db      142 ProCysThrGluIleValTyrProGluAsnGlyValLeuCysGlyArgAsnValMetAla 161
QY      487 CTGAGTGCCTCCCGCGGTGCGCCACTGACCTGCGCACCTGCGACACCACTACCGTCCAAG 546
Db      162 LeuLysVal---GlyAspArgArgLeuIleCysHisLeuTyrThrSerTyrArgSerLys 180
QY      547 AAGCCCGCTCCGCGCTGAGATGCGCGCTTCCACTTCGAGGACCAACCGATCGAGATC 606
Db      181 LysAlaValArgAlaLeuThrMetProGlyPheHisPheThrAspIleArgLeuGln--- 199
QY      607 CTGAGGAGGTGGAGAGGCAAGTGTACAAAGCAGTACGAGCGCGCGCTGGCGCGTAC 666
Db      200 MetProArgLysLysLysAspGlyTyrPheGluLeuTyrGluAlaSerValAlaArgTyr 219
QY      667 TGCAGCGCGCGCGCTCCAAG 687
Db      220 SerAsp---LeuProGluLys 225

RESULT 6
Q95W85
ID Q95W85 PRELIMINARY; PRT; 227 AA.
AC Q95W85;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein-like protein.
OS Condylactis gigantea (giant anemone).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynantheae; Actiniidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RX Matz M.V.; Lukyanov S.A.;
RT "Diversity and evolution of GFP-like fluorescent proteins.";
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037777; AAK71343.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25384 MW; D3C6B02F490F3D21 CRC64;

```


Alignment Scores:
 Pred. No.: 8,32e-42 Length: 227
 Score: 747.50 Matches: 143
 Percent Similarity: 76.65% Conservative: 31
 Best Local Similarity: 63.00% Mismatches: 50
 Query Match: 57.50% Indels: 3
 DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x Q8M4U5 (1-227)

QY 7 GCTCCCTCTGACCGACCATGCCCTTCAGGACCCATCGAGGACCCGCTGAACGGC 66
 |||||
 Db 2 AlaGlyLeuLeuLysGluSerMetArgIleLysIleTyrMetGluGlyThrValasnGly 21
 |||||
 QY 67 CACTACTTCAAGTACCGGACGAGGCGAGGCGAACCCTTCGAGGCGACCCAGGAGATG 126
 |||||
 Db 22 TyrHisPheLysCysGluGlyGluGlyAspGlyAsnProPheGluGlyThrGlnAsnMet 41
 |||||
 QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTCGCTTCACATCTCCATCTCTCCACCTCC 186
 |||||
 Db 42 ArgIleAsgValThrGluGlyAlaProLeuProPheAlaPheAspIleLeuSerProCys 61
 |||||
 QY 187 TGCATGTAGGCTTCAAGCGCTTCATCAAGTACGTGTCGGCATCTCCGACTACTTCAAG 246
 |||||
 Db 62 CysAlaTyrGlySerLysThrPheIleLysHisThrSerGlyIleProAspTyrPheLys 81
 |||||
 QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCCACCTACGAGGACGCGGCTTC 306
 |||||
 Db 82 ArgSerPheProGluGlyPheThrTrpGluGlyThrThrIleTyrGluAspGlyArgVal 101
 |||||
 QY 307 CTGACCGCCACCGAGCACCTCCCTGAGCGGAGCTGCTGGTGTTACAAAGGTGAAGATC 366
 |||||
 Db 102 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnProIleTyrLysValLysVal 121
 |||||
 QY 367 CTGGCAACAATCTCCCGCGCGCGCTGATGACAGACCAAGCGCGCGCTCGGAG 426
 |||||
 Db 122 LeuGlyThrAsnPheProAlaLaserProValMetLysLysValSerGlyGlyTrpGlu 141
 |||||
 QY 427 CCCTCCACCGAGATCGTGTACAGGTGTGACGGCTGCTGGCGGCCAGTCCCTGATGGCC 486
 |||||
 Db 142 ProSerThrGluIleValTyrGlnAspAsnGlyValLeuArgGlyArgAsnValMetAla 161
 |||||
 QY 487 CTGGAGTCCCGCGCGCTGCGCACCTGACCTGCCACCTGCACACCACTACCGCTCCAAG 546
 |||||
 Db 162 LeuLysValSerGlyArgProProLeuIleCysHisLeuHisSerThrTyrArgSerLys 181
 |||||
 QY 547 AAGCCCGCTCCGCTGAAGATCCCGCTTCCACTTCGAGGACCAACCCATCGAGATC 606
 |||||
 Db 182 Lys---AlaCysAlaLeuThrMetProGlyPheHisPheAlaAspLeuArgIleGln--- 199
 |||||
 QY 607 CTGGAGGAGTGGAGAGGCAAGTGTACAGCAGTACGAGGCGCGCGCTGGCGCTAC 666
 |||||
 Db 200 MetProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 219
 |||||
 QY 667 TCGAGCGCGCGCGCTCCCAAG 687
 |||||
 Db 220 SerAsp---ValProGluLys 225

RESULT 7
 Q8ISF8 PRELIMINARY; PRT; 231 AA.
 ID Q8ISF8;
 AC Q8ISF8;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Red fluorescent protein Pp611
 OS Paracyonias actinotoxoides (Sea anemone) (Entamea quadricolor).
 OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Actinaria;
 OC Nyantheae; Actiniidae; Entacmaea.
 OX NCBI_TaxID=6118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22199846; PubMed=12185250;

Wiedemann J., Schenk A., Roecker C., Girod A., Spindler K.-D.,
 RA Nienhaus G.U.;
 RT "A far-red fluorescent protein with fast maturation and reduced
 RT oligomerization tendency from Entamea quadricolor (Anthozoa,
 RT Actinaria)."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:11646-11651 (2002).
 DR EMBL; AY130757; AAN05449.1; -;
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 231 AA; 26053 MW; 86467A7B5D6DD60 CRC64;

Alignment Scores:
 Pred. No.: 2,84e-41 Length: 231
 Score: 739.50 Matches: 139
 Percent Similarity: 72.37% Conservative: 26
 Best Local Similarity: 60.96% Mismatches: 62
 Query Match: 56.88% Indels: 1
 DB: 5 Gaps: 1

US-10-081-864-21 (1-707) x Q8ISF8 (1-231)

QY 10 TCCCTGTGTGACCGAGACCATGCCCTTCAGGACCCATCGAGGACCCGCTGAACGGCCAC 69
 |||||
 Db 3 SerLeuIleLysGluAsnMetArgMetValValMetGluGlySerValasnGlyTyr 22
 |||||
 QY 70 TACTTCAAGTGCACCGCAAGGCGGAGGCGAACCCTTCGAGGCGACCCAGGAGATGAAG 129
 |||||
 Db 23 GlnPheLysCysThrGlyGluGlyAspGlyAsnProTyrMetGlyThrGlnThrMetArg 42
 |||||
 QY 130 ATCAGGTGTATCGAGGCGCGCCCTCGCTTCCACATCTGCTCCACCTCTGCTGC 189
 |||||
 Db 43 IleLysValValGluGlyGlyProLeuProPheAlaPheAspIleLeuAlaThrSerPhe 62
 |||||
 QY 190 ATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGCATCCCGACTCTTCAAGCAG 249
 |||||
 Db 63 MetTyrGlySerLysThrPheIleLysHisThrLysGlyIleProAspPheLysGln 82
 |||||
 QY 250 TCCCTCCCGAGGCTTCACCTCGGAGCGCACCCACCTACGAGGACCGCGCTTCCTG 309
 |||||
 Db 83 SerPheProGluGlyPheThrTrpGluArgValThrArgTyrGluAspGlyGlyValPhe 102
 |||||
 QY 310 ACCGCCACACGACACCTCCCTCGGACGGGAGCTGCTGTGTACAGGTGAAGATCCTG 369
 |||||
 Db 103 ThrValMetGlnAspThrSerLeuGluAspGlyCysLeuValTyrHisAlaLysValThr 122
 |||||
 QY 370 GGCAACAACATTCGCCCGCGCGCTGATGCAGAACCAAGCGCGCTCGGAGGCC 429
 |||||
 Db 123 GlyValasnPheProSerAsnGlyAlaValMetGlnLysLysThrLysGlyTyrDgluPro 142
 |||||
 QY 430 TCCACCGAGATCGTGTACGAGGTGGAGCGGCTGCTCGCGCGCGCTGCTGATGGCCCTG 489
 |||||
 Db 143 AsnThrGluMetLeuTyrProAlaAspGlyGlyLeuArgGlyTyrSerGlnMetAlaLeu 162
 |||||
 QY 490 GAGTGGCCCGCGCTGCGCACCTGACCTGCACCTGCACACCACTACCGCTCCAAGAAG 549
 |||||
 Db 163 AsnValaspGlyGlyGlyTyrLeuSerCysSerPheGluThrThrTyrArgSerLysLys 182
 |||||
 QY 550 CCCGCTCCCGCTGAAGATGCCCGCTTCCATTCAGGACCAACCGCATTCGAGATCCTG 609
 |||||
 Db 183 ThrValGluAsnPheLysMetProGlyPheHisPheValAspHisArgLeuGluArgLeu 202
 |||||
 QY 610 GAGGAGTGGAGAGGCAAGTGTCTACAGCAGTACGAGGCGCGCGCTGCGCGCGCTGCTG 669
 |||||
 Db 203 GluGluSerAspLysGluMetPheValValGluHisGluHisAlaValAlaLysPheCys 222
 |||||
 QY 670 GACCGCGCGCGCTCCCAAGCTGGGC 693
 |||||
 Db 223 Asp---LeuProSerLysLeuGly 229

RESULT 8

Q9GP16
ID Q9GP16 PRELIMINARY; PRT; 228 AA.
AC Q9GP16
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein asFP499 (Green fluorescent protein as(s)FP499).
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actinidae; Anemonia.
OX NCBI_TaxID=6108;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20570437; PubMed=11121018;
RA Wiedenmann J., Elke C., Spindler K.-D., Funke W.;
RT "Cracks in the beta-can: Fluorescent proteins from Anemonia sulcata (Anthozoa, Actiniaria).";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14091-14096(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Leutenegger A., Wiedenmann J.;
RT "Photobiology of Anemonia sulcata.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF322221; AAG41205.1; -;
DR EMBL; AF545827; AAN52735.1; -;
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 228 AA; 25369 MW; 89CEFFB7E1ECDE73 CRC64;

Alignment Scores:
Pred. No.: 1.89e-34 Length: 228
Score: 637.00 Matches: 126
Percent Similarity: 67.54% Conservative: 28
Best Local Similarity: 55.26% Mismatches: 70
Query Match: 49.00% Indels: 4
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q9GP16 (1-228)

QY 16 CTGACCGAGACCATCCCTTCAGACACCATCGAGGACCGTGAACGGCCACTACTTC 75
Db 5 IleLysGluThrMetArgValGlnLeuSerMetGluGlySerValAsnTyrHisAlaPhe 24
QY 76 AAGTGCACCGGCAAGGCGAGGGCAACCCCTCGAGGACCCAGGAGATGAAGATCGAG 135
Db 25 LysCysThrGlyLysGlyGluGlyLysProTyrGluGlyThrGlnSerLeuAsnIleThr 44
QY 136 GTGATCAGGGGGGCCCCCTGCGCTTCGCATCCATCCATCGTCCACCTCTCTCATGTAC 195
Db 45 IleThrGluGlyGlyProLeuProPheAlaPheAspIleLeuSerHisAlaPheGlnTyr 64
QY 196 GGTCTCAAGCCCTTCATCAAGTACGTCTCGGATCCCGACTACTTCAAGCAGTCCCTC 255
Db 65 GlyIleLysValPheAlaLysTyrProLysGluIleProAspPheLysGlnSerLeu 84
QY 256 CCGAGGGCTTCACCTGGGAGCGACACCCACCATCAGGAGACCGCGGCTTCCTGACCGCC 315
Db 85 ProGlyGlyPheSerTrpGluArgValSerThrTyrGluAspGlyGlyValLeuSerAla 104
QY 316 CACCAGACACCTCCCTGACGCGGACTCGCTGTGTACAGGTGAAGATCCTGGGCAAC 375
Db 105 ThrGlnGluThrSerLeuGlnGlyAspCysIleLeuValLysValLysValLeuGlyThr 124
QY 376 AACTTCCCGCCGACCGCCCTGTGATGCAGAACAGCGCGCGCTGGGAGCCCTCCACC 435
Db 125 AsnPheProAlaAsnGlyProValMetGlnLysThrCysGlyTrpGluProSerThr 144
QY 436 GAGATCGTGTACGAGTGACGCGCTGTGCGGGCCAGTCCCTGTGATGCGCTGAGATGTC 495

Db 145 GluThrValIleProArgAspGlyGlyLeuLeuArgAspThrProAlaLeuMetLeu 164
QY 496 CCGGGGGTGGCCACCTGACCTGCCACCTGCACACCATCAGCTCCCAAGAACGCCGCC 555
Db 165 AlaAspGlyGlyHisLeuSerCysPheMetGluThrThrTyrLysSerLysLys----- 182
QY 556 TCGCCCTGAAGATGCCCGCTTCCCTTCGAGGACCCCGCATCGAGATCCTGGAGGAG 615
Db 183 ---GluValLysLeuProGluLeuHisPheHisLeuArgMetGluLysLeuAsnIle 201
QY 616 GTGGAGAAGCGCAAGTGCTACAGAGCAGTACGAGGCGCGCTGCGCGCTTACTGCGACGCC 675
Db 202 SerAspAspTrpLysThrValGluGlnHisGluSerValValAlaSerTyr---SerGln 220
QY 676 GCCCCTCCCAAGCTGGGCCACCAAC 699
Db 221 ValProSerLysLeuGlyHisAsn 228
RESULT 9
Q720W9
ID Q720W9 PRELIMINARY; PRT; 225 AA.
AC Q720W9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Red fluorescent protein.
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mc1.
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V., Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181552; AAO61598.1; -;
SQ SEQUENCE 225 AA; 25822 MW; 830871F6936953E2 CRC64;

Alignment Scores:
Pred. No.: 2e-29 Length: 225
Score: 561.50 Matches: 109
Percent Similarity: 63.64% Conservative: 31
Best Local Similarity: 49.55% Mismatches: 75
Query Match: 43.19% Indels: 5
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q720W9 (1-225)

QY 10 TCCCTGTGTCAGGAGACCATCGCTTCAGGACCATCGAGGACCGTGAACGGCCAC 69
Db 2 SerValIleLysSerValMetLysIleLysLeuArgMetGluGlySerValAsnGlyHis 21
QY 70 TACTTCAAGTCCACCGCAAGGCGGAGGCAACCCCTCGAGGACCCAGGAGATGAAG 129
Db 22 AsnPheValIleValGlyGluGlyLysProTyrGluGlyThrGlnSerMetAsp 41
QY 130 ATCAGAGTGTACGAGGCGGCCCTCGCTTCGCTTCACATCTCCATCTCCACCTCTCTGC 189
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleMetThrThrValPhe 61
QY 190 ATGTACGGCTCCAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAGCAG 249
Db 62 HisTyrGlyAsnArgValPheAlaLysTyrProLysHisIleProAspTyrPheLysGln 81
QY 250 TCCCTCCCGGGCTTCACCTGGGAGCGGCACACCATCAGGACCGCGGCTTCCTG 309
Db 82 MetPheProGluGlyTyrSerTrpGluArgSerMetAsnPheGluAspGlyIleCys 101
QY 310 ACCGCCACGAGGACCATCTCCCTGGAGCGGACTGCTGTGTACAGGTGAAGATCCTG 369

```

Db      102 ThrAlaArgAsnGluIleThrMetGluGlyAspCysPheAsnLysValArgPheAsp 121
QY      370 GGCAACAACATTCCTCCGCGGAGCGGCTGATGAGAAACAGCGCGCGCTGGAGCCC 429
Db      122 GlyValAsnPheProPAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY      430 TCCACCGAGATCGTGTCAGAGGTGGAGCGGCTGCTGGCGGCGGCGGCTGATGGCCCTG 489
Db      142 SerThrGluLysMetTyrValArgAspGlyValLeuThrGlyAspIleAsnMetAlaLeu 161
QY      490 GAGTGGCCCGCGCTGCCACCTGACCTGCACCTGCACACCACTACCGCTCCAGAGAG 549
Db      162 LeuLeuGluGlyGlyHisTyrArgCysAspPheArgThrThrArgAlaLysLys 181
QY      550 CCGCCCTCCGCTCAGAGATCGCGGCTTCCACTTCGAGGACCAACCGCATCGAGATCCTG 609
Db      182 -----LysGlyValLysLeuProAspTyrHisPheValAspHisSerIleGluIleLeu 199
QY      610 -----GAGGAGGTGGAGAGGCAAGTGTACAGCAGTACGAGGCGCGCTGGGC 660
Db      200 ArgHisAspLysGluTyrThrGluValLysLeuTyrGluHisAlaGluAlaHisSerGly 219

RESULT 10
Q8T5E7 PRELIMINARY; PRT; 229 AA.
ID AC Q8T5E7
AC Q8T5E7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CgIFP-g.
OS Condylactis gigantea (Giant anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nyanthesea; Actiniidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21927629; PubMed=11929936;
RA Labas V.A., Gurskaya N.G., Yamushevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT "Diversity and evolution of the green fluorescent protein family.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261 (2002).
DR GO; AY037776; AAK71342.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 229 AA; 25731 MW; A5E4E1CCFE8E85B4 CRC64;

Alignment Scores:
Pred. No.: 2.34e-29 Length: 229
Score: 560.50 Matches: 113
Percent Similarity: 60.79% Conservative: 25
Best Local Similarity: 49.78% Mismatches: 84
Query Match: 43.12% Indels: 5
DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x Q8T5E7 (1-229)

QY 16 CTGACCGAGACCATGCCCTTCAGGACCACTGAGGGGACCGTGAACGGCGCTACTTC 75
Db 5 IleLysGluThrMetArgSerLysValTyrMetGluGlyAspValAsnAsnHisAlaPhe 24
QY 76 AAGTGCAACCGGCAAGGCGGAGGCAACCCCTCGAGGGACCCAGGAGATGAAGATCGAG 135
Db 25 LysCysThrAlaValGlyGluGlyLysProTyrLysGlySerGlnAspLeuThrIleThr 44
QY 136 GTGATCAGGCGGCGCCCTTCGCTTCGCTTCACATCCTGTCCACCTCCCTGCATGTAC 195
Db 45 ValThrGluGlyGlyProLeuProPheAlaPheAspIleLeuSerHisAlaPheGlnTyr 64
QY 196 GGCTCCCAAGGCTTCATCAAGTAGTCGTCCGGCATCCCGGACTCTTCAGCAGTCCCTC 255

```

```

Db      65 GlyAsnLysValPheThrAspTyrProAspAspIleProAspPhePheLysGlnSerLeu 84
QY      256 CCCGAGGGCTTCACCTGGGAGCGCACCACTACGAGGAGCGGGCTTCTCTGACCGCC 315
Db      85 SerAspGlyPheThrTrpArgValSer***Tyr***GlyGlyValLeuThrVal 104
QY      316 CACGAGGACACCTCCCTGGAGCGGCTGCTGCTGTACAGGTGAAGTGAAGATCCTGGGCAAC 375
Db      105 ThrGlnAspThrSerLeuLysGlyAspCysIleIleCysAsnIleLysValHisGlyThr 124
QY      376 AACTTCCCGCGGAGCGGCGGCTGATGAGAACAGGCGCGCGCTGGAGGCC---TCC 432
Db      125 AsnPheProGluAsnGlyProValMetGlnAsnLysThrAspGlyTrpGluProSerSer 144
QY      433 ACCGAGATCGTGTACGAGGTGACGCGCTGCTGCGCGGCCAGTCCCTGATGCGCCCTGGAG 492
Db      145 ThrGluThrValIleProGlnAspGlyGlyIleValAlaAlaArgSerProAlaLeuArg 164
QY      493 TGCCCGCGGCGTGGCGCACCTGACCTGCCACCTGCACACCACTACCGCTCCAGAGAGCCC 552
Db      165 LeuArgAspLysGlyHisLeuIleCysHisMetGluThrThrTyr-----LysPro 181
QY      553 GCCTCCGCGCTCAGAGATGCCCGGCTTCACCTTCGAGGACCAACCGCATCCGATCCTGGAG 612
Db      182 AsnLysGluValLysLeuProGluLeuHisPheHisLeuArgMetGluLysLeuSer 201
QY      613 GAGGTGGAGAGGGCAAGTGTCTACAGCAGTACAGGCGCGCGCTGGCGGCTACTGCGAC 672
Db      202 ValSerAspAspGlyLysThrIleLysGlnHisGluTyrValValAlaSerTyr---Ser 220
QY      673 GCGCGCCCTCCCAAGCTGGC 693
Db      221 LysValProSerLysIleGly 227

RESULT 11
Q8MU48 PRELIMINARY; PRT; 224 AA.
ID AC Q8MU48
AC Q8MU48;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein-like protein.
OS Montastraea annularis (boulder star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=48500;
RN [1]
RP SEQUENCE FROM N.A.
RA Matz M.V., Lukyanov S.A.;
RT "Diversity and evolution of GFP-like fluorescent proteins.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037766; AAK71332.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 224 AA; 25452 MW; 1C4D85F87D6315EC CRC64;

Alignment Scores:
Pred. No.: 2.94e-29 Length: 224
Score: 559.00 Matches: 108
Percent Similarity: 63.18% Conservative: 31
Best Local Similarity: 49.03% Mismatches: 75
Query Match: 43.00% Indels: 6
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q8MU48 (1-224)

QY 10 TCCTCGCTGACGAGACCATGCCCTTCAGGACCACTGAGGGGACCGTGAACGGCCAC 69
Db 2 SerMetIleLysProGluMetLysIleLysMetArgMetAspGlyAlaValAsnGlyHis 21

```



```
SQ SEQUENCE 227 AA; 26035 MW; 2C41B38DCB84279C CRC64;
Alignment Scores:
Pred. No.: 3,17e-29 Length: 227
Score: 558.50 Matches: 112
Percent Similarity: 63.68% Conservative: 30
Best Local Similarity: 50.22% Mismatches: 72
Query Match: 42.96% Indels: 9
DB: 5 Gaps: 4

US-10-081-864-21 (1-707) x Q95VT0 (1-227)
QY 10 TCCTCTGCTGACGAGACCGTTCGCTTCAGGACACCATCGAGGCGACCGTGAACGGCCAC 69
Db 2 SerValIleLysProAspMetLysLysLeuArgMetGluGlyAlaValAsnGlyHis 21
QY 70 TACTTCAAGTGCACCGGCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATGAAG 129
Db 22 LysPheValIleGluGlyAspGlyLysGlyLysProPheGluGlyThrGlnSerMetAsp 41
QY 130 ATCAGAGTGATCGAGGCGGCGCCCTCGCTTCGCTTCACATCTGCTCCACCTCTGTC 189
Db 42 LeuThrValIleGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrThrValPhe 61
QY 190 ATGTAGGCTCCAGGCTTCATCAAGTACGTCTCGGATCCCGACTCCGACTACTTCAAGCAG 249
Db 62 AspTyrGlyAsnArgValPheAlaLysTyrProGlnAspIleThrAspTyrPheLysGln 81
QY 250 TCCTCTCGGAGGCTTCACCTGGAGCGGCGCCCTGATGAGGCGGCGGCTTCCTG 309
Db 82 ThrPheProGluGlyTyrPheTyrPgluArgSerMetThrTyrGluAspGlnGlyLysCys 101
QY 310 ACCGCCACACAGAC---ACCTCCCTGGACGGC-----GACTGCTGTGTACAGGTG 360
Db 102 IleAlaThrAsnAspIleThrMetMetLysGlyValAspAspCysPheValTyrLysIle 121
QY 361 AAGATCTCGGCGCAACACTTCCCGCGGCGGCGCTGATGAGGCGGCGGCGGCGGCG 420
Db 122 ArgPheAspGlyValAsnPheProAlaAsnGlyProValMetGlnArgLysThrLeuLys 141
QY 421 TGGAGCGCTTCCACGAGATCGTGTACGAGGTGGACGGCGTGTGCGGCGGCGGCGGCTG 480
Db 142 TrpGluProSerThrGluIleMetTyrAlaArgAspGlyValLeuLysGlyAspValAsn 161
QY 481 ATGGCTCTGAGTGGCGGCGGCTGCGACCTGACCTGCGACCTGCGACACCGCTACCGC 540
Db 162 MetAlaLeuLeuLeuGluGlyGlyGlyHisTyrArgCysAspPheLysThrThrThrLys 181
QY 541 TCCAAAGACCGCGCTCCGCGCTGAGATCGCGGCTTCCACTTCGAGGACCAACCGCATC 600
Db 182 AlaLysLys-----ValValArgLeuProAspTyrHisPheValAspHisArgIle 198
QY 601 GAGATCTCTG-----GAGGAGGTGGAGAGGCGGAGTGTCTACAGCAGTACGAGGCG 651
Db 199 GluIleValSerHisAspLysAspTyrAsnLysValLysLeuTyrGluHisAlaGluAla 218
QY 652 GCGTGGCG 660
Db 219 *****Gly 221

RESULT 14
Q8T5F1 PRELIMINARY; PRT; 225 AA.
AC Q8T5F1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE McavFP 7.5.
OS Montastrea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastrea.
OC NCB1_TaxID=63558;
RX NCB1_1
```

```
SEQUENCE FROM N.A.
MEDLINE=21927629; PubMed=11929996;
RA Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT "Diversity and evolution of the green fluorescent protein family.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR EMBL; AY037770; AAK71336.1;
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green fl protein; 1.
SQ SEQUENCE 225 AA; 25866 MW; 820C89437F8BDB32 CRC64;
Alignment Scores:
Pred. No.: 6.83e-29 Length: 225
Score: 553.50 Matches: 108
Percent Similarity: 63.18% Conservative: 31
Best Local Similarity: 49.09% Mismatches: 76
Query Match: 42.58% Indels: 5
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q8T5F1 (1-225)
QY 10 TCCTCTGCTGACGAGACCGTTCGCTTCAGGACACCATCGAGGCGACCGTGAACGGCCAC 69
Db 2 SerValIleLysSerValMetLysLysLeuArgMetGluGlySerValAsnGlyHis 21
QY 70 TACTTCAAGTGCACCGGCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATGAAG 129
Db 22 AsnPheValIleValGlyGluGlyLysProTyrGluGlyThrGlnSerMetAsp 41
QY 130 ATCAGAGTGATCGAGGCGGCGCCCTGCTTCGCTTCACATCTGCTCCACCTCTGTC 189
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleMetThrThrValPhe 61
QY 190 ATGTAGGCTCCAGGCTTCATCAAGTACGTCTCGGACTCCCGCATCCCGACTACTTCAAGCAG 249
Db 62 HisTyrGlyAsnArgValPheAlaLysTyrProLysHisIleProAspTyrPheLysGln 81
QY 250 TCCTCTCGGAGGCTTCACCTGGAGCGGCGACCCACCTACGAGGCGGCGGCTTCCTG 309
Db 82 MetPheProGluGluTyrSerTrpGluArgSerMetAsnPheGluGlyGlyLysCys 101
QY 310 ACCGCCACCGGCGACACCTCCCTGCGGCGGCGCTGCTGTGTACAGGTGAAGATCTCTG 369
Db 102 ThrAlaArgAsnGluIleThrMetGluGlyAspCysPhePheAsnLysValArgPheAsp 121
QY 370 GGCAACAACTTCCCGCGGCGGCGGCTGTGTATGAGAAACAAGCGCGCGCTGGAGGCC 429
Db 122 GlyValAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY 430 TCCAGGAGATCTGTACGAGGTGAGCGGCTGCTGCGGCGGCGGCTGCTGATGCGGCTG 489
Db 142 SerThrGluLysMetTyrValArgAspGlyValLeuThrGlyAspIleAsnMetAlaLeu 161
QY 490 GAGTGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCGGCTGCGGCGGCGGCTGCGGCG 549
Db 162 LeuLeuGluGlyGlyGlyHisTyrArgCysAspPheArgThrThrTyrArgAlaLysLys 181
QY 550 CCGGCTCGGCGGCTGAGTACCGGCTTCACCTTCGAGGCGGCGGCGGCTGAGTACCTG 609
Db 182 -----LysGlyValLysLeuProAspTyrHisPheGluAspHisSerIleGluLeu 199
QY 610 -----GAGGAGGTGGAGGCGGCAAGTGTCTACAGCAGTACGAGGCGGCGGCGGCG 660
Db 200 ArgHisAspLysGluTyrThrGluValLysLeuTyrGluHisAlaGluAlaHisSerGly 219

RESULT 15
Q8T6U0 PRELIMINARY; PRT; 236 AA.
ID Q8T6U0
AC Q8T6U0;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: August 3, 2004, 10:03:29 ; Search time 4729.5 Seconds
(without alignments)
6479.234 Million cell updates/sec

Title: US-10-081-864-21
Perfect score: 707
Sequence: 1 ggatcgctcctctgtgac.....ctgggccaactgaagctt 707

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.ste.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vit.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	707	100.0	707	6	AX824739 Sequence
2	672	95.0	699	12	AY233273 Synthetic
3	649.2	91.8	654	6	AX824741 Sequence
4	440.2	62.3	696	6	AX686890 Sequence
5	440.2	62.3	696	6	AX824727 Sequence
6	436.8	61.8	955	3	AF246709 Anemolia
7	428.8	60.7	699	3	AF222222 Anemolia
8	417.8	59.1	687	6	AX527898 Sequence
9	417.8	59.1	687	6	AX527900 Sequence
10	417.8	59.1	1396	6	AX527902 Sequence
11	417.8	59.1	1424	6	AX527904 Sequence
12	315	44.6	675	6	AX824731 Sequence
13	315	44.6	1050	6	AX666133 Sequence
14	313.4	44.3	678	6	AX370408 Sequence
15	313.4	44.3	678	6	AX824732 Sequence
16	313.4	44.3	681	12	AF506025 Synthetic
17	313.4	44.3	1395	12	AF506026 Synthetic
18	313.4	44.3	10141	12	AY342347 Red H-Pel
19	313.4	44.3	10276	12	AY342348 Red H-Pel
20	312.6	44.2	678	12	AF506027 Synthetic
21	310.2	43.9	678	6	AX370404 Sequence
22	310.2	43.9	678	6	AX824725 Sequence
23	308.6	43.6	678	6	AX370406 Sequence
24	308.6	43.6	4692	6	AX463702 Sequence
25	308.6	43.6	6893	6	AX823860 Sequence
26	308.6	43.6	9320	6	AX663075 Sequence
27	286.2	40.5	681	6	AX233584 Sequence
28	286.2	40.5	713	6	AX233627 Sequence
29	284	40.2	835	3	AF363775 Condylact
30	284	40.2	835	3	AF363776 Heteracti
31	282.4	39.9	681	6	AX573118 Sequence
32	278.8	39.4	684	3	AF383155 Condylact
33	272.4	38.5	835	3	AY037777 Condylact
34	271.8	38.4	678	6	AX824729 Sequence
35	267.6	37.9	681	6	AX527894 Sequence
36	267.6	37.9	681	6	AX527896 Sequence
37	266	37.6	908	3	AF363776 Heteracti
38	264.4	37.4	684	6	AX527910 Sequence
39	264.4	37.4	910	6	AX527888 Sequence
40	264.4	37.4	910	6	AX527914 Sequence
41	263	37.2	696	3	AY130757 Entacmaea
42	262.8	37.2	684	6	AX527892 Sequence
43	261.6	37.0	690	6	AX824737 Sequence
44	261.2	36.9	908	6	AX527890 Sequence
45	254	35.9	680	6	AX527912 Sequence

ALIGNMENTS

RESULT 1
AX824739
LOCUS AX824739
DEFINITION Sequence 21 from Patent WO02068459.
ACCESSION AX824739
VERSION AX824739.1
KEYWORDS GI:39750599
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS
TITLE
JOURNAL
FEATURES

707 bp DNA linear PAT 11-DEC-2003

Handwritten signature/initials

Db 364 CTGTGTAATATTTCTCTGCTGATGGCCCCGCTGATGACAGAAACAGAGGAGATGGAG 423
 QY 427 CCTTCCACCGAGATCGTGTACAGGTGACCGCGTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 424 CCATCCACCGAGATAGTTTATGAAGTTGACGGTGTCTGCGTGGACAGTCTTTGATGGCC 483
 QY 487 CTGGAGTSCCGCGGCGTCCACCTGACCTGCGCACCTGACACACACCTACCGCTCCAAG 546
 Db 484 CTTAAGTSCCGTGGTGTGCTATCTGACTTGGCATCTCCATATCTTACAGTCCAAA 543
 QY 547 AAGCCCGCTCGCCCTGAAGATGCCCGCTTCCATCTCGAGGACCAACCGCATCGAGATC 606
 Db 544 AAACAGCTGTGCTTGAAGATGCCAGATTTTCAATTTGAAGATCATGCGATCGAGATA 603
 QY 607 CTGGAGGAGTGGAGAGGCAAGTGTACAGAGTACAGGCGCGCGTGGGCGGTAC 666
 Db 604 ATGGAGGAGTGGAGAGGCAAGTGTATTAACAGTACGAAGCAGCAGTGGGCGGTAC 663
 QY 667 TGGCAGCGCGCCCTCCCAAGCTGGGCCACAAC 699
 Db 664 TGTGATGCTGTCTCCATCCCAAGCTTGGACATAC 696

RESULT 5
AX824727

LOCUS AX824727 696 bp DNA linear PAT 11-DEC-2003
 DEFINITION Sequence 9 from Patent WO02068459.
 ACCESSION AX824727
 VERSION AX824727.1 GI:39750592

KEYWORDS Anemonia sulcata (snake-locks sea anemone)
 SOURCE Anemonia sulcata
 ORGANISM Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;

REFERENCE 1 Nynantheae; Actiniidae; Anemonia.

AUTHORS Non aggregating fluorescent proteins and methods for using the same
 TITLE Patent: WO 02068459-A 9 06-SEP-2002;
 JOURNAL Location/Qualifiers

FEATURES
 source

1..696
 /organism="Anemonia sulcata"
 /mol_type="unassigned DNA"
 /db_xref="taxon:6108"

ORIGIN

Query Match 62.3%; Score 440.2; DB 6; Length 696;
 Best Local Similarity 77.2%; Pred. No. 9e-46;
 Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 7 GCCTCCCTGTCGCGGAGACCATGCCCTTACAGACCAATCGAGGCGCACCGTGAACGGC 66
 Db 4 GCCTCCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTTGAAGGCGCGTTAATGGC 63
 QY 67 CACTACTTCAAGTGCACCGGAGGCGGCAACCCCTCGAGGCGCACCGAGGATG 126
 Db 64 CACTACTTCAAGTGTACAGAAAGAGAGGCGGCAACCATTTGAGGGTACGAGGAATG 123
 QY 127 AAGATCGAGGTGTCGAGGCGGCGCCCTGCGCTTCCGCTTCCACATCTGTCCTCC 186
 Db 124 AAGATAGAGGTTCATCGAAGGAGGTCTCATTTGCCATTTTGCCTTCCACATTTTGTCAAGAGT 183
 QY 187 TGCATGACCGCTCCAGGCGCTTCATCAAGTACGTGTCGCGCATCCCGACTACTTCAAG 246
 Db 184 TGTATGACGCTAGTAGAGGCTTCATCAAGTATGTGTCAGGAATCTTGTACTACTTCAAG 243
 QY 247 CAGTCCCTCCCGAGGCTTCACTGGAGCGCACCAACCTTACGAGGACGCGCGCTTC 306
 Db 244 CAGTCTTCTCCCTGAAGGTTTACTTGGGAAGAAGCAACCAACCTTACGAGGATGAGGCTTT 303
 QY 307 CTGACCGCCCAACGAGACCTCCCTGGAGCGGCTGCTGCTGTGATGAGGTGATGATC 366
 Db 304 CTTACAGCTCATCAGGACACAAGCCTAGATGGAGATTCCTCTGTTTACAGGTCAGATT 363

QY 367 CTGGGCAACAACCTTCCCGCCGACGCGCCCTGATGACAGAAACAGGCGCGCTGGGAG 426
 Db 364 CTGTGTAATATTTCTCTGCTGATGGCCCCGCTGATGACAGAAACAGGAGATGGGAG 423
 QY 427 CCTTCCACCGAGATCGTGTACAGGTGAGCGCGTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 424 CCATCCACCGAGATAGTTTATGAAGTTGACGGTGTCTCGTGGACAGTCTTTGATGGCC 483
 QY 487 CTGGAGTSCCGCGCGGTGCGCACCTGACCTGCCACCTGACACACCTACCGTCCCAAG 546
 Db 484 CTTAAGTSCCGTGGTGTGCTATCTGACTTGGCATCTCCATATCTTACAGTCCAAA 543
 QY 547 AAGCCCGCTCGCCCTGAAGATGCCCGCTTCCATCTCGAGGACCAACCGCATCGAGATC 606
 Db 544 AAACAGCTGTGCTTGAAGATGCCAGATTTTCAATTTGAAGATCATGCGATCGAGATA 603
 QY 607 CTGGAGGAGTGGAGAGGCAAGTGTACAGAGTACAGGCGCGCGTGGGCGGTAC 666
 Db 604 ATGGAGGAGTGGAGAGGCAAGTGTATTAACAGTACGAAGCAGCAGTGGGCGGTAC 663
 QY 667 TGGCAGCGCGCCCTCCCAAGCTGGGCCACAAC 699
 Db 664 TGTGATGCTGTCTCCATCCCAAGCTTGGACATAC 696

RESULT 6

AF246709
 LOCUS AF246709 955 bp mRNA linear INV 29-AUG-2000
 DEFINITION Anemonia sulcata GFP-like chromoprotein FP595 mRNA, complete cds.
 ACCESSION AF246709
 VERSION AF246709.1 GI:9937257

KEYWORDS Anemonia sulcata (snake-locks sea anemone)
 SOURCE Anemonia sulcata
 ORGANISM Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;

REFERENCE 1 (bases 1 to 955)
 Nynantheae; Actiniidae; Anemonia.

AUTHORS Lukyanov, A.F., Fradkov, A.F., Gurskaya, N.G., Matz, M.V., Labas, Y.A., Savitsky, A.P., Markelov, M.L., Zaraisky, A.G., Zhao, X., Fang, Y., Tan, W. and Lukyanov, S.A.

Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog

J. Biol. Chem. 275 (34), 25879-25882 (2000)

REFERENCE 2 (bases 1 to 955)
 10852900

AUTHORS Matz, M.V., Fradkov, A.F., Lukyanov, K.A., Gurskaya, N.G. and Lukyanov, S.A.

Direct Submission
 Submitted (17-MAR-2000) Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117871, Russia

FEATURES

source
 1..955
 /organism="Anemonia sulcata"
 /mol_type="mRNA"
 /db_xref="taxon:6108"

69..767
 /note="purple chromoprotein; similar to green fluorescent protein"

CDS
 /product="GFP-like chromoprotein FP595"

/protein_id="AAG02385.1"

/db_xref="GI:9937258"

/translation="MASFLKMTMPFKTTIETGTVNGHYFKTGKGNPFGTQBMKIE
 VIEGGLPFAFHILSTSCMTGSKTFIKYSGIDYFKQSPFEGFTWERTTITDGGFL
 TAHQDTSLDGLVYKVLGNPPADGPNKAGNWEPAIVIEVDGVLGQSILM
 ALKCPGGRHLTCHLHTYRSKKPASALKMPGFHRIIMEVEKGCYKQYEAAY
 GRYCDAAPSKLGHN"

ORIGIN

Query Match 61.8%; Score 436.8; DB 3; Length 955;
 Best Local Similarity 76.7%; Pred. No. 2.2e-45;
 Matches 534; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACATCCCTTCAGGACCACTCGAGGACCCGCTGAACGGC 66
 Db |||||
 72 GCTTCCTTTTAAAGAGACTATCCCTTTAAGACGACCACTGAAGGACGGTTAATGGC 131
 QY 67 CACTACTTCAAGTACACCGGCAAGGCGGCAACCCCTCGAGGACCCAGGAGATG 126
 Db |||||
 132 CACTACTTCAAGTACAGGAAGAGAGAGGCAACCCCTCGAGGACCCAGGAGATG 191
 QY 127 AAGATCAGAGTATCGAGGCGGCGCCCTGCTTCGCTTCACATCTCTGCTCCACTCC 186
 Db |||||
 192 AAGATAGAGGTATCGAGGAGGTCATTCCTTCACATCTCTGCTTCACATCTTGTCAACGAT 251
 QY 187 TGCATGACGGCTCCAAAGGCTTCATCAAGTACGCTGTCGGGACCTCCCGACTACTTCAAG 246
 Db |||||
 252 TGTATGACGCTAGTAAGACCTTCATCAAGTATGTCAGGATTCCTGACTACTTCAAG 311
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGGACACACCTTACAGAGCGGCGCTTC 306
 Db |||||
 312 CAGTCTTTCCTGAGGCTTACTTGGGAAGAACCAACCTTACGAGGATGGAGGCTTT 371
 QY 307 CTGACCGCCACACGAGACACCTCCCTGGACGGGACTGCTGCTGTACAGGTTGAAGATC 366
 Db |||||
 372 CTTACAGCTCATCAGGACACAAAGCTAGATGGAGATTGCTCTCTTACAGGTCAGATT 431
 QY 367 CTGGGCAACACTTCCCGCGAGCGCCCTGATGAGAGAAAGCGCGCGCTGGAG 426
 Db |||||
 432 CTTGGTAAATTTCTGCTGATGGCCCGCTGATGAGAACCAAGCAGGAAGTGGAG 491
 QY 427 CCCTCCACCGAGATCGTGTACGAGTGGACGGGCTGCTGCGCGGCGAGTCCCTGATGGC 486
 Db |||||
 492 CCAGCCACGAGATAGTTATGAAGTGAAGGCTGCTGCTGCGAGAGCTTTTGTGGCC 551
 QY 487 CTGAGAGTCCCGCGGCTGCGACCTGACCTGCACTGCAACACCTACCGCTCCAG 546
 Db |||||
 552 CTTAAGTGCCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
 QY 547 AAGCCCGCTCCGCGCTGAGAGTCCCGGCTTCCACTTCGAGGACCAACGCTGAGATC 606
 Db |||||
 612 AAACAGCTAGTGCCTGAGAGTCCAGGATTTTCATTTGAAGATCATCGATCGAGATA 671
 QY 607 CTGAGAGTGGAGAGGCAAGTGTCAACAGTACGAGCGCGCGCTGGCGGCTAC 666
 Db |||||
 672 ATGAGGAGTGTGAGAGGCAAGTGTCTAAGACAGTACGAGCAGTGGGCGAGGTAC 731
 QY 667 TGCACGCGCGCGCTCCAGCTGGGCGCAACTGA 702
 Db |||||
 732 TGTGATGCTGCTCATCCAGCTTGGACATACTAA 767

RESULT 7
 AF322222
 LOCUS
 DEFINITION Anemonia sulcata nonfluorescent red protein asCP562 mRNA, complete cds.
 ACCESSION AF322222
 VERSION AF322222.2 GI:18999345
 KEYWORDS
 SOURCE
 ORGANISM
 Anemonia sulcata (snake-locks sea anemone)
 Anemonia sulcata
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Nynantheae; Actiniidae; Anemonia.
 1 (bases 1 to 699)
 Wiedemann, J., Elke, C., Spindler, K.D. and Funke, W.
 Cracks in the beta-can: fluorescent proteins from Anemonia sulcata
 (Anthozoa, Actiniaria)
 Proc. Natl. Acad. Sci. U.S.A. 97 (26), 14091-14096 (2000)
 JOURNAL MEDLINE
 PUBLISHED 20570437
 1121018
 2 (bases 1 to 699)
 Wiedemann, J., Elke, C., Spindler, K.-D. and Funke, W.
 Direct Submission
 REFERENCE
 TITLE
 Submitted (18-NOV-2000) Allgemeine Zoologie und Endokrinologie,

Universitaet Ulm, Albert-Einstein-Allee 11, Ulm 89069, Germany
 3 (bases 1 to 699)
 Wiedemann, J., Elke, C., Spindler, K.-D., Funke, W., Sundin, B. and Jach, G.
 Direct Submission
 Submitted (28-FEB-2002) Allgemeine Zoologie und Endokrinologie,
 Universitaet Ulm, Albert-Einstein-Allee 11, Ulm 89069, Germany
 Sequence update by submitter
 On Feb 28, 2002 this sequence version replaced gi:11890752.
 Location/Qualifiers
 1..699
 /organism="Anemonia sulcata"
 /mol_type="mRNA"
 /db_xref="taxon:6108"
 1..699
 /notes="similar to GFP"
 /codon_start=1
 /product="nonfluorescent red protein asCP562"
 /protein_id="AA01206.2"
 /db_xref="GI:18999346"
 /translation="MASFLKKTMPFKTTIEGTVNGHYFKCTCKGNGPFGTQEMKIE
 VIRGGPLPAFHILSTCMYGSKTFKVVSGIPDYFKQSPFGFTWERTTTTVDGGL
 TAQDTSLDGDCIVKVKILGNFPADGPMONKAEKPEPAETILYEVDGVLRGSLM
 ALKCPGGRHLTCHLHSTYRSKXPSALNKPFGPHGDRHRIEMEEVEKCKYQYAAV
 ARYCDAAAPSKLGH"

ORIGIN
 Query Match 60.7%; Score 428.8; DB 3; Length 699;
 Best Local Similarity 76.0%; Pred. No. 2.4e-44;
 Matches 529; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACATCCCTTCAGGACCACTCGAGGACCCGCTGAACGGC 66
 Db |||||
 4 GCTTCCTTTTAAAGAGACTATCCCTTTAAGACGACCACTGAAGGACGGTTAATGGC 63
 QY 67 CACTACTTCAAGTACACCGGCAAGGCGGCAACCCCTCGAGGACCCAGGAGATG 126
 Db |||||
 64 CACTACTTCAAGTACAGGAAGAGAGAGGCAACCCCTCGAGGACCCAGGAGATG 123
 QY 127 AAGATCAGTATCGAGGCGGCGCCCTGCTTCGCTTCACATCTCTGCTCCACTCC 186
 Db |||||
 124 AAGATAGAGGTATCGAGGAGGTCATTCCTTCCACTTTCCTTTCCTTTCCTTTCCT 183
 QY 187 TGCATGACGGCTCCAAAGGCTTCATCAAGTACGCTGTCGGGACCAACGCTGAGATC 246
 Db |||||
 184 TGTATGACGCTAGTAAGACCTTCATCAAGTACGCTGTCAGGATTCAGACTACTTCAAG 243
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGGACCAACCTTACGAGCGCGCGCTTC 306
 Db |||||
 244 CAGTCTTTCCTTGAAGGTTTACTTGGGAAGAACCAACCTTACGAGGATGGAGCTTT 303
 QY 307 CTGACCGCGCCACCGAGACACCTCCCTGGACGGGACTGCTGCTGTACAGGTTGAAGATC 366
 Db |||||
 304 CTTACAGCTCATCAGGACACAAAGCTTAGATGGAGATTGCTGCTTTTACAGGTCAGATT 363
 QY 367 CTGGGCAACAACTTCCCGCGGACCGCCCTGCTGATGATGAGAAACAAAGCGCGCTGGAG 426
 Db |||||
 364 CTTGGTAAATTTTCTGCTGATGGCCCGCTGATGAGAACCAAGCAGGAAGTGGAG 423
 QY 427 CCCTCCACCGAGATCGTGTACGAGGTCGCGCTGCTGCGCGGCGAGTCCCTGATGGC 486
 Db |||||
 424 CCAGCCACCGAGATCTTATGAGTTGAGGTCCTGCTGCTGGAGAGCTTTTGTATGGCC 483
 QY 487 CTGAGTGTCCCGCGGCTGCGACCTGACCTGCCACTGTGACACCAACCTTACCGCTCCAG 546
 Db |||||
 484 CTTAAGTGCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
 QY 547 AAGCCCGCTCCGCGCTGAGAGTGGCCGCTTCCACTTCGAGGACCAACGCTCCAGATC 606
 Db |||||
 544 AAACAGCTAGTGCCTTGAAGATGCGAGATTTCATTTTGGAGATCATCGCATCAGATA 603
 QY 607 CTGAGAGTGGAGAGGCAAGTGTCTAAGACAGTACGAGCGCGCGCTGGCGGCTAC 666
 Db |||||

Db 604 ATGGAGGAAGTTGAGAAAGGCAAGTGCTATTAACAGTACGAAGCAGCAGTGGCCAGGTAC 663
QY 667 TGGCAGCGCGCCCTCCCAAGCTGGCCACAACTGA 702
Db 664 TGTGATGCTGCTCCATCCAAAGCTTGGACATCACTAA 699

RESULT 8
AX527898
LOCUS AX527898 687 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 11 from Patent WO0230965.
ACCESSION AX527898
VERSION AX527898.1 GI:25172348
KEYWORDS Heteractis crispa (leathery sea anemone);
ORGANISM Heteractis crispa
SOURCE Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
REFERENCE 1
AUTHORS Lukyanov, S.A., Pradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 11 18-APR-2002;
Clontech Laboratories Inc. (US)
FEATURES
source
1. .687
/organism="Heteractis crispa"
/mol_type="unassigned DNA"
/db_xref="taxon:175771"

ORIGIN
Query Match 59.1%; Score 417.8; DB 6; Length 687;
Best Local Similarity 79.1%; Pred. No. 5.5e-43;
Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CCTGCTGACCGAGACCATGCCCTTCAGGACCAACCCCTCGAGGACCCGCTGAAGGCACTA 71
Db 12 CCTGCTGACCGAGAGATGCGCATCAAGATGTATGAGGGGACCCGTGAACGGCCACTA 71
QY 72 CTTCAAGTGCACCGGAGGCGGAGGCAACCCCTCGAGGACCCGAGAGATGAAGAT 131
Db 72 CTTCAAGTGCAGGCGGAGGCGGAGGCAACCCCTCGAGGACCCGAGAGATGAAGAT 131
QY 132 CGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCCTGTCACCTCTCTGCAT 191
Db 132 CCAGTGCAGGCGGCGCCCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 191
QY 192 GTACGGGTCCAAAGGCTTCATCAAGTACGTGTCGGGATCCCGGATCCCGGATCCCGG 251
Db 192 GTACGGGTCCAAAGGCTTCATCAAGTACGTGTCGGGATCCCGGATCCCGGATCCCGG 251
QY 252 CTTCCCGGAGGCTTCACCTGGAGAGACCCACCTACGAGGCGGCGGATCCCTGAC 311
Db 252 CTTCCCGGAGGCTTCACCTGGAGAGACCCACCTACGAGGCGGCGGATCCCTGAC 311
QY 312 CGCCACACAGGACACCTCCCTGGAGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431
Db 312 CACCACTTCCCGGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 431
QY 432 CACCGAGATGCTGTAAGAGTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491
Db 432 CACCGAGTGTGTATCCCGGAGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 491
QY 492 GTGCCCCGGGCTGCGCCACCTGACCTGCGGAGGACCCGAGGACCCGAGGACCCGAGG 551
Db 492 G----GTGGGGGACCGGACCTGATCTGCGGAGGACCCGAGGACCCGAGGACCCGAGG 548
QY 552 CGCCTCCGCGCTGAAGATCCCGGCTTCACCTTCGAGGACCCGAGGACCCGAGGACCCGAG 611
Db 549 CGTGGCGGCGCTGACCATGCGCGGCTTCACCTTCGAGGACCCGAGGACCCGAGGACCCGAG 608

QY 612 GGAGGTGGAGAGGCAAGTGTCTACAAGCAGTACGAGCCGCGCTGGGCGGCTACTCTGCGA 671
Db 609 GAA---GAAGAAGGACGAGTACTTTCGAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCGA 665

QY 672 C 672
Db 666 C 666

RESULT 9
AX527900
LOCUS AX527900 687 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 13 from Patent WO0230965.
ACCESSION AX527900
VERSION AX527900.1 GI:25172349
KEYWORDS Heteractis crispa (leathery sea anemone);
ORGANISM Heteractis crispa
SOURCE Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
REFERENCE 1
AUTHORS Lukyanov, S.A., Pradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 13 18-APR-2002;
Clontech Laboratories Inc. (US)
FEATURES
source
1. .687
/organism="Heteractis crispa"
/mol_type="unassigned DNA"
/db_xref="taxon:175771"

ORIGIN
Query Match 59.1%; Score 417.8; DB 6; Length 687;
Best Local Similarity 79.1%; Pred. No. 5.5e-43;
Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CCTGCTGACCGAGACCATGCCCTTCAGGACCAACCCCTCGAGGACCCGCTGAAGGCACTA 71
Db 12 CCTGCTGAGGAGAGATGCGCATCAAGATGTATGAGGGGACCCGTGAACGGCCACTA 71
QY 72 CTTCAAGTGCACCGGAGGCGGAGGCAACCCCTCGAGGACCCGAGAGATGAAGAT 131
Db 72 CTTCAAGTGCAGGCGGAGGCGGAGGCAACCCCTTCGCGGACCCGAGAGATGCGGAT 131
QY 132 CGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCCTGTCACCTCTCTGCAT 191
Db 132 CCAGTGCAGGCGGCGCCCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 191
QY 192 GTACGGGTCCAAAGGCTTCATCAAGTACGTGTCGGGATCCCGGATCCCGGATCCCGG 251
Db 192 GTACGGGTCCAAAGGCTTCATCAAGTACGTGTCGGGATCCCGGATCCCGGATCCCGG 251
QY 252 CTTCCCGGAGGCTTCACCTGGAGAGACCCACCTACGAGGCGGCGGATCCCTGAC 311
Db 252 CTTCCCGGAGGCTTCACCTGGAGAGACCCACCTACGAGGCGGCGGATCCCTGAC 311
QY 312 CGCCACACAGGACACCTCCCTGGAGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371
Db 312 CGCCACACAGGACACCTCCCTGGAGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371
QY 372 CAACACTTCCCGGAGGCGGCGGCTGATGTCAGAAACAGGCGGCGGCTGGGAGGCCCTC 431
Db 372 CACCACTTCCCGGAGGCGGCGGCTGATGTCAGAAACAGGAGGCGGCGGCTGGGAGGCC 431
QY 432 CACCGAGATGCTGTAAGAGTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491
Db 432 CACCGAGTGTGTATCCCGGAGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 491
QY 492 GTGCCCCGGGCTGCGCCACCTGACCTGCGGAGGACCCGAGGACCCGAGGACCCGAGG 551
Db 492 G----GTGGGGGACCGGACCTGATCTGCGGAGGACCCGAGGACCCGAGGACCCGAGG 548
QY 552 CGCCTCCGCGCTGAAGATCCCGGCTTCACCTTCGAGGACCCGAGGACCCGAGGACCCGAG 611
Db 492 G----GTGGGGGACCGGACCTGATCTGCGGAGGACCCGAGGACCCGAGGACCCGAGG 548

QY 552 CGCCTCCGCTGAAGATCCCGGCTTCCACTTCGAGGACACCGCATCGAGATCCTGGA 611
|
Db 549 CGTGGGGCCCTGACCATGCCGGCTTCCACTTCACCGACATCCCGCTGCAGATCCTGG 608
|
QY 612 GGAGGTGGAGAGGCAAGTGTACAAGCAGTACAGGCGCGCTGGGCGGTACTTGCGA 671
|
Db 609 GAA---GGAGAGGACGAGTACTTCGAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCA 665
|
QY 672 C 672
|
Db 666 C 666

RESULT 10
AX527902
LOCUS AX527902 1396 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 15 from Patent WO0230965.
ACCESSION AX527902
VERSION AX527902.1 GI:25172350
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Lukyanov,S.A., Fradkov,A.F., Lukyanov,K.A. and Gurskaya,N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 15 18-APR-2002;
Clontech Laboratories Inc. (US)

FEATURES
source
1..1396
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="fusion construct"

ORIGIN

Query Match 59.1%; Score 417.8; DB 6; Length 1396;
Best Local Similarity 79.1%; Pred. No. 4.5e-43;
Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTTCGAGGACACCCAGGAGATGAAGAT 131
|
Db 25 CTTCAAGTGCACCGGCAAGGGGAGGGGACGCAACCCCTTCGCGCGCACCCAGAGCATGCGGAT 144
|
QY 132 CGAGGTGATCGAGGG 191
|
Db 145 CCACGTGACGAGGG 204
|
QY 192 GTACGGCTCCAAAGGCTTTCATCAAGTACGTGTCCGGCATCCCGCATCTTCAAGCAGTC 251
|
Db 205 GTAGGACGAGGACCTTCGTGCACACACACCGCGAGATCCCGCATCTTCAAGCAGAG 264
|
QY 252 CTTCCCGAGGGGTTTCACTGGAGGGGACACACCTACCTACGAGGAGGGGGGCTTCTGAC 311
|
Db 265 CTTCCCGAGGGGTTTCACTGGAGGAGAACCCACCTACGAGGAGGGGGGCTTCTGAC 324
|
QY 312 CGCCCAACGAGGACACCTCCTCGAGCGGCGACTGCTGTGTACAAGGTGAAGATCCTGGG 371
|
Db 325 CGCCCAACGAGGACACCAAGCTTGGAGGGCACTGCTGTGTACAAGGTGAAGATCCTGGG 384
|
QY 372 CAACAACTTCCCGCGCGCGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 431
|
Db 385 CACCAACTTCCCGCGCGCGCGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 444
|
QY 432 CACGAGATCTGT 491
|
Db 445 CACCGAGGTGGTGTACCCCGAGAACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 504
|
QY 492 GTGCCCCGGGGT 551
|

Db 505 G---GTGGGACCGCGGCTGTCTCCACCATCACAGTACCGGAGCAGAGGC 561
|
QY 552 CGCCTCCGCTGAAGATCCCGGCTTCCACTTCGAGGACACCGCATCGAGATCCTGGA 611
|
Db 562 CGTGGGGCCCTGACCATGCCGGCTTCCACTTCACCGACATCCGGCTGCAGATCCTGCG 621
|
QY 612 GGAGGTGGAGAGGCAAGTGTACAAGCAGTACAGGCGCGCTGGGCGGTACTTGCGA 671
|
Db 622 GAA---GGAGAGGACGAGTACTTCGAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCA 678
|
QY 672 C 672
|
Db 679 C 679

RESULT 11
AX527904
LOCUS AX527904 1424 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 17 from Patent WO0230965.
ACCESSION AX527904
VERSION AX527904.1 GI:25172351
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Lukyanov,S.A., Fradkov,A.F., Lukyanov,K.A. and Gurskaya,N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 17 18-APR-2002;
Clontech Laboratories Inc. (US)

FEATURES
source
1..1424
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="fusion construct"

ORIGIN

Query Match 59.1%; Score 417.8; DB 6; Length 1424;
Best Local Similarity 79.1%; Pred. No. 4.5e-43;
Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTTCGAGGACACCCAGGAGATGAAGAT 131
|
Db 25 CTTCAAGTGCACCGGCAAGGGGAGGGGACGCAACCCCTTCGCGCGCACCCAGAGCATGCGGAT 144
|
QY 72 CTTCAAGTGCACCGGCAAGGGGAGGGGACGCAACCCCTTCGAGGACACCCAGGAGATGAAGAT 131
|
Db 85 CTTCAAGTGCACCGGCAAGGGGAGGGGACGCAACCCCTTCGCGCGCACCCAGAGCATGCGGAT 144
|
QY 132 CGAGGTGATCGAGGG 191
|
Db 145 CCACGTGACGAGGG 204
|
QY 192 GTAGGGCTCCAAAGGCTTTCATCAAGTACGTGTCCGGCATCCCGCATCTTCAAGCAGTC 251
|
Db 205 GTAGGGGAGGGGTTTCACTGGAGGGGACACCTACCTACGAGGAGGGGGGCTTCTGAC 264
|
QY 252 CTTCCCGAGGGGTTTCACTGGAGGAGAACCCACCTACGAGGAGGGGGGCTTCTGAC 311
|
Db 265 CTTCCCGAGGGGTTTCACTGGAGGAGAACCCACCTACGAGGAGGGGGGCTTCTGAC 324
|
QY 312 CGCCCAACGAGGACACCTCCTCGAGCGGCGACTGCTGTGTACAAGGTGAAGATCCTGGG 371
|
Db 325 CGCCCAACGAGGACACCAAGCTTGGAGGGCACTGCTGTGTACAAGGTGAAGATCCTGGG 384
|
QY 372 CAACAACTTCCCGCGCGCGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 431
|
Db 385 CACCAACTTCCCGCGCGCGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 444
|
QY 432 CACGAGATCTGT 491
|

Db 445 CACGAGGTGGTGTACCCCGAGAACGGCGTGTGTGGGCGCGAAACGTGATGGCCCTGAA 504
Qy 492 GTGCCCCGGGGTGGCCACCTGACCTGCACTGCAACACACACCTACCGCTTCAAGAGGC 551
Db 505 G---GTGGGGACCGGGCGGTGATGTCGCCACCACTACACAGCTACCGAGCAAGAAGGC 561
Qy 552 CGCTCCCGCTGAAGATGCCGGCTTCCACTTCGAGGACCCGCGATCGAGATCTCTGA 611
Db 562 CGTCGGGGCTTACCATGCGCGCTTCCACTTACCGCATCCGGCTGCAATGCTGG 621
Qy 612 GGAGGTGAGAAGGCAAGTGTCTACAGCACTACGAGCGCGCGTGGCGGCTTACTGCGA 671
Db 622 GAA---GGAGAAGGACAGTACTTCGAGCTGTACGAGGCCAGCGTGGCGCGGTACAGCGA 678
Qy 672 C 672
Db 679 C 679

RESULT 12
AX824731
LOCUS AX824731 675 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 13 from Patent WO02068459.
ACCESSION AX824731
VERSION AX824731.1 GI:39750594
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Non aggregating fluorescent proteins and methods for using the same
TITLE Patent: WO 02068459-A 13 06-SEP-2002;
JOURNAL Location/Qualifiers
FEATURES
source
1..675
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="non-aggregating mutant"

ORIGIN
Query Match 44.6%; Score 315; DB 6; Length 675;
Best Local Similarity 68.5%; Pred. No. 3.2e-30;
Matches 454; Conservative 0; Mismatches 200; Indels 9; Gaps 1;
Qy 4 TCCGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCACTCGAGGGCACCGTGAAC 63
Db 10 TCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGCGCACCGTGAAC 69
Qy 64 GGCCACTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTTCAGAGGGCACCCAGGAG 123
Db 70 GGCCACAGTTCGAGATCGAGGGCGAGGGCGAGGGCGGCGCCCTTACGAGGGCCCAACACC 129
Qy 124 ATGAAGATCGAGGTGATCGAGGGCGGCGCCCTTCGCTTTCACATCTCTGTCACAC 183
Db 130 GTGAAGTGAAGGTGACCAAGGGCGGCGCCCTTCGCTTTCGCTGAGACATCTCTGTCGCC 189
Qy 184 TCCTGATGATGAGCTCCAGGGCTTCATCAAGTACGTGTGCGGCATTCCTCGGACTACTTC 243
Db 190 CAGTTCAGTACGGCTCCAGGTGTACGTGAAGCACCCCGCGGACATCCCGGACTACAAG 249
Qy 244 AAGCAGTCCCTCCCGAGGGCTTCACTGGAGCGGACACCACTACGAGGACGGCGG 303
Db 250 AAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGCTGATGAATTCGAGGACGGCGG 309
Qy 304 TTCTGTACCGGCCACCAAGGACCTCCCTGAGCGGACATGCTGCTGTGTACAAAGTGAAG 363
Db 310 GTGGCGACCGTGAACCAAGGACTCTCCCTCGAGGACGGCTTCTATCTACAAAGTGAAG 369
Qy 364 ATCTGTGGCAACACTTCCCGCGGCGGCGGCTGTGATGAGAACAGGCGGCGGCTGG 423
Db 370 TTATCGGGGTGAACCTTCCCTTCGAGCGGCGGCGGCTGTGATGAGAACAGGACCATGGGCTGG 429

Qy 424 GAGCCCTCCACCGAGATCGTGTACGAGTGTGACGCGCTGTGTGCGCGCCAGTCCCTCATG 483
Db 430 GAGGCTTCCACCGAGCGCTGTACCCCGGACGCGCTGTGTGAAGGCGAGACCCACAAG 489
Qy 484 GCCCTGAGTGTCCCGCGGTCGCTGACCTGACCTGCCACCTGACACACCACTACCGTCC 543
Db 490 GCCCTGAAGCTGAAGGACGCGCGCACTACCTGTGTGAGTTCAAGTCCATCTACATGCC 549
Qy 544 AAGAGCCCGCTCCGCGCTGAAGATGCCGGCTTCCACTTCGAGGACCAACCGATCGAG 603
Db 550 AAGAAGCCCG-----TGCAGTCCCGGCTTACTACTACGTGGACGCAAGCTGGAC 600
Qy 604 ATCTGAGGAGGTGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCTGGGCGCG 663
Db 601 ATCACTCCCAACAGGAGTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCG 660
Qy 664 TAC 666
Db 661 CAC 663

RESULT 13
AX666133
LOCUS AX666133 1050 bp DNA linear PAT 26-MAR-2003
DEFINITION Sequence 7 from Patent WO02060941.
ACCESSION AX666133
VERSION AX666133.1 GI:29290961
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified.

REFERENCE 1
AUTHORS Zhao, M., Xu, M., Jiang, P. and Yang, M.
TITLE Fluorescent proteins
JOURNAL Patent: WO 02060941-A 7 08-AUG-2002;
Anticancer, Inc. (US)
FEATURES
Location/Qualifiers
source
1..1050
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
/note="Coral"

ORIGIN
Query Match 44.6%; Score 315; DB 6; Length 1050;
Best Local Similarity 68.5%; Pred. No. 2.9e-30;
Matches 454; Conservative 0; Mismatches 200; Indels 9; Gaps 1;
Qy 4 TCCGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCACTCGAGGGCACCGTGAAC 63
Db 298 TCCGAGAACGTCATCACCGAGTTTCATGGCTTCAAGTGGCGATGGAGGCGCACCGTGAAC 357
Qy 64 GGCCACTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTTCGAGGGCACCCAGGAG 123
Db 358 GGCCACAGTTCGAGATCGAGGGCGAGGGCGAGGGCGGCGCCCTTACGAGGGCCCAACACC 417
Qy 124 ATGAAGATCGAGGTGATCGAGGGCGGCGCCCTTCGCTTTCACATCTCTGTCACAC 183
Db 418 GTGAAGCTGAAGGTGACCAAGGGCGGCGCCCTTCGCTTTCGCTTGGGACATCTCTGTCGCC 477
Qy 184 TCCTGATGATGAGCTCCAGGCTTCAAGGCTTCATCAAGTACGTGTGCGGCATTCCTCGGACTACTTC 243
Db 478 CAGTTCAGTACGGCTCCAGGTGTACGTGAAGCACCCCGCGGACATCCCGGACTACAAG 537
Qy 244 AAGCAGTCCCTCCCGAGGGCTTCACTGGAGCGGACCACTACGAGGACGGCGG 303
Db 538 AAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGCTGATGAACCTTCGAGGACGGCGG 597
Qy 304 TTCCTGACCGGCCACCAAGGACCTCCCTGAGCGGCGACTGCCTGTGTGTACAAAGTGAAG 363
Db 598 GTGGCGACCGTGAACCAAGGACTCTCCCTCGAGGACGGCTGCTTCACTACAAAGTGAAG 657
Qy 364 ATCTGTGGCAACACTTCCCGCGGCGGCGGCTGTGATGAGAACAGGCGGCGGCTGG 423

Db 658 TTTCATCGGCGTGAACCTCCCTCCGACGCGCCCGTGTATCAGAGAAGACCATGGGCTGG 717
Qy 424 GAGCCCTCCACCGAGATCGTACGAGGTGAGCGGCTGCTGGCGGCCAGTCCCTGTATG 483
Db 718 GAGGCTTCACCGAGCGCTGTATCCCGCGACGGCTGCTGAAGGCGAGACCCACAAG 777
Qy 484 GCGCTGAGTGGCCCGCGGCTGCGCCACCTGACCTGCGCCACCTGACACCACTACCGCTCC 543
Db 778 GCGCTGAAGTGAAGACGCGCGCCACTACTCTGCTGAGTTCAGTTCATCTACATGGCC 837
Qy 544 AAGAAGCGCGCTCCCGCTCGAGATCGCGGCTTCCACTTCGAGGACACCGCATCGAG 603
Db 838 AAGAAGCGCG-----TGCAGCTGCCCGGCTACTACTAGTGAAGCGCAAGCTGAC 888
Qy 604 ATCTCGAGAGGTGAGAGGCAAGTGTCTAAGCAGTACGAGCGCGCTGAGCGCGC 663
Db 889 ATCACTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 948
Qy 664 TAC 666
Db 949 CAC 951

RESULT 14
AX370408
LOCUS AX370408 578 bp DNA linear PAT 16-FEB-2002
DEFINITION Sequence 5 from Patent WO0196373.
ACCESSION AX370408
VERSION AX370408.1 GI:18857492
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
1 artificial sequences.
REFERENCE 1
AUTHORS Pradkov,A.F. and Tersikh,A.
TITLE Fluorescent timer proteins and methods for their use
JOURNAL Patent: WO 0196373-A 5 20-DEC-2001;
Clontech Laboratories Inc. (US)
FEATURES
source
1..678
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="variant of sequence from Discosoma sp."

ORIGIN
Query Match 44.3%; Score 313.4; DB 6; Length 678;
Best Local Similarity 68.3%; Pred. No. 5.1e-30;
Matches 453; Conservative 0; Mismatches 201; Indels 9; Gaps 1;
Qy 4 TCCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAAC 63
Db 10 TCCGAGAACGTATCATCCGAGTTCATCGCTTCAAGTGGCATGAGGCGACCGTGAAC 69
Qy 64 GGCCTACTTCAAGTGCACCGGCAAGGCGAGGCGAACCCCTTCGAGGCGACCCAGGAG 123
Db 70 GGCACAGATTGAGATCGAGGCGAGGCGAGGCGCGCCCTTACGAGGCGCCACAACACC 129
Qy 124 ATGAGATCGAGTATCGAGGCGCGCCCTTCGCTTCCCATCTTCATCTTCCTCCACC 183
Db 130 GTGAAGTGAAGTGAACAAGGCGCGCCCTTCGCTTCCGCTGGGACATCTCTGTCCCC 189
Qy 184 TCCTGCTATGACGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTC 243
Db 190 CAGTTCAGTACGCTCCAAAGTGTAGTGAAGCACCCCGGACATCCCGACTACAAG 249
Qy 244 AAGCAGTCCCTCCCGAGGGCTTCACTGGGAGGCGACCACTACGAGGAGCGGGC 303
Db 250 AAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGTATGAATTCGAGGACGGCGC 309
Qy 304 TTCCTGACCGCCACCAAGACACTCTCCTGAGCGGCACTGCCTGGTGTATCAAGTGAAG 363

Db 310 GTGCGGACCGTGAACCCAGGACTCCTCCCTCGAGGACGGCTTCTATCTATAAGGTGAAG 369
Qy 364 ATCTGGGCAACAATTTCCCGCGGACGCGCCCGTGTATGAGAACAAAGGCGCGCGCTGG 423
Db 370 TTCTATCGGCGTGAACCTTCCCTCCGACGCGCCCGTGTATGAGAAAGAACCATGGGCTGG 429
Qy 424 GAGCCCTCCACCGAGATCGTGTAGAGGTGAGCGGCTGCTGCGCGCCAGTCCCTGTATG 483
Db 430 GAGGCTTCACCGAGCGCTGTATCCCGCGGACGCGCTGCTGAAGGCGGAGATCCACAAG 489
Qy 484 GCGCTGAGTGGCCCGCGGCTGCGCCACCTGACCTGCGCCACCTGCGACACCACTACCGTCC 543
Db 490 GCGCTGAAGTGAAGGACGCGCGCCACTACTCTGCTGAGTTCAGTTCATCTATACGGCC 549
Qy 544 AAGAAGCGCGCTCCCGCTCGAGATCGCGGCTTCCACTTCGAGGACCAACCGCATCGAG 603
Db 550 AAGAAGCGCG-----TGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
Qy 604 ATCTCGAGGAGGTGAGAGGCAAGTGTCTAAGCAGTACGAGGCGCGCTGGGCGCGC 663
Db 601 ATCACTCCCAACAGGAGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
Qy 664 TAC 666
Db 661 CAC 663

RESULT 15
AX824732 678 bp DNA linear PAT 11-DEC-2003
LOCUS AX824732
DEFINITION Sequence 14 from Patent WO02068459.
ACCESSION AX824732
VERSION AX824732.1 GI:39750595
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
1 artificial sequences.
REFERENCE 1
AUTHORS Non aggregating fluorescent proteins and methods for using the same
TITLE Patent: WO 02068459-A 14 06-SEP-2002;
JOURNAL Location/Qualifiers
FEATURES
source
1..678
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="non-aggregating mutant"

ORIGIN
Query Match 44.3%; Score 313.4; DB 6; Length 678;
Best Local Similarity 68.3%; Pred. No. 5.1e-30;
Matches 453; Conservative 0; Mismatches 201; Indels 9; Gaps 1;
Qy 4 TCCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAAC 63
Db 10 TCCGAGAACGTATCATCCGAGTTCATCGCTTCAAGTGGCATGAGGCGACCGTGAAC 69
Qy 64 GGCCTACTTCAAGTGCACCGGCAAGGCGAGGCGAACCCCTTCGAGGCGACCCAGGAG 123
Db 70 GGCACAGATTGAGATCGAGGCGAGGCGAGGCGCGCCCTTACGAGGCGCCACAACACC 129
Qy 124 ATGAGATCGAGTATCGAGGCGCGCCCTTCGCTTCCCATCTTCATCTTCCTCCACC 183
Db 130 GTGAAGTGAAGTGAACAAGGCGCGCCCTTCGCTTCCGCTGGGACATCTCTGTCCCC 189
Qy 184 TCCTGCTATGACGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTC 243
Db 190 CAGTTCAGTACGCTCCAAAGTGTAGTGAAGCACCCCGGACATCCCGACTACAAG 249
Qy 244 AAGCAGTCCCTCCCGAGGGCTTCACTGGGAGGCGACCACTACGAGGAGCGGGC 303
Db 250 AAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGTATGAATTCGAGGACGGCGC 309

QY 304 TTCCTGAGCGCCACCAGGACACCTCCCTGGACGGCGACTGCTGGTGTACAAGGTGAAG 363
Db |||||
QY 310 GTGGGACCGTGAACCCAGGACTCTCTCTGAGGACGGCTGCTCACTACAAGGTGAAG 369
Db |||||
QY 364 ATCTGGGACACAACTTCCCGCCGACGGCCCGGTGATGACAGAAACAAGGCCGGCCGCTGG 423
Db |||||
QY 370 TTCATCGCGGTGAATTCCTCTCCGACGGCCCGGTGATGACAGAAACAAGGCCGGCTGG 429
Db |||||
QY 424 GAGCCCTCCACCGAGATCGTGTACGAGGTGACGGCGTGTGCGGCGCCAGTCCCTGATG 483
Db |||||
QY 430 GAGGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAGGGGAGATCCACAAG 489
Db |||||
QY 484 GCCCTGGAGTGCCTCGCGGTGCGCACCTGACCTGCCACCTGCACACCACTACCGTCC 543
Db |||||
QY 490 GCCCTGAAGCTGAAGGACGGCGGCACTACCTGGTGGAGTTCAAGTCCATCATGGCC 549
Db |||||
QY 544 AAGAGCCCGCTCCGCGCTGAAGATCCCGGCTTCCACTTCGAGGACCAACCGCATCGAG 603
Db |||||
QY 550 AAGAGCCCG-----TGCAGTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
Db |||||
QY 604 ATCTGTGAGGAGTGGAGAGGCAAGTGTACAAGCAGTACGAGGCGCGCTGGGCGCGC 663
Db |||||
QY 664 TAC 666
Db |||||
QY 661 CAC 663
Db |||||

Search completed: August 3, 2004, 14:04:19
Job time : 4732.5 secs

Db 327 GGAGATCCGACGACATCAACCTGATCGAGCAAGTTCTGTACCGCTGGAGTACAA 386
Qy 369 GGGCAACAATTCCTCCGCGACGCGCCGCTGATCGAGCAAGCCGCGCTGGAGCC 428
Db 387 GGGCAGCAATTCCTCCGCGACGCGCGCTGATCGAGCAAGCCATCTCTGGCATCGAGCC 446
Qy 429 CTCACCCAGATCTGTACGAGGTGGAGCGGCTGCTGCGCGGCGAGTCCCTGATGGCCCT 488
Db 447 CAGCTTCAGGCCATGTATACATGAACAACGCGCTGCTGTGGGCGAGGTATCTCTGTGTA 506
Qy 489 GGAGTCCCGCGCTGCGCACCTGACCTGCCACTGACACACCTACCGCTCCAAAGAA 548
Db 507 CAAGCTGAACAGCGCAGTACTACAGCTGCCACATGAGAC-----CTGTGAGAG 560
Qy 549 GCGCCCTCCGCTCGAGATGCGCGCTTCCACTTTCGAGGACACCGCATCGAGATCCT 608
Db 561 CAAGGCGCTGTGAAGGAGTTCCCTCTCATCCACTTCATCCAGCAGCGCTGGAGA---A 617
Qy 609 GGAGGAGTGGAGAGGCAAGTGTCAACAGCAGTACGAGCGCGCTGGGCC 661
Db 618 GACCTAGTGGAGGCGCGCTTCTGTGGAGCAGCAGAGCCGCTCGCCC 670

RESULT 2

US-09-459-956-7
; Sequence 7, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1230-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavularia sp
US-09-459-956-7

Query Match 23.3%; Score 165; DB 4; Length 801;
Best Local Similarity 54.9%; Pred. No. 1.7e-24;
Matches 352; Conservative 0; Mismatches 280; Indels 9; Gaps 1;
Qy 27 CATGCCCTTCAGGACCACTCGAGGCGACCGTGAACGCGCACTACTCAAGTGCACCGG 85
Db 147 CATGAAGATTAACTGAAGTGAAGGAATGTAACCGGATCGTCTTTGTGTATCGAAG 206
Qy 87 CAAGGGCGAGGGCAACCCCTCGAGGGCACCAGGAGATGAAGATCGAGGTGATCGAGGG 146
Db 207 AGNAGGAAGAAAGCCCTTACGATGGGACACACACTTTAAACCTGGAAGTGAAGGAAG 266
Qy 147 CGGCCCCCTGCGCTTCCACATCTGTCCACCTCTGATGATACCGCTCCAGGCTCCAGGC 206
Db 267 TGGCGCTCTGCTTTCTTACGATATCTTGTCAACCGGCTTCAGTACCGAAACAGAGC 326
Qy 207 CTTCATCAAGTACGTGTCCGCACTCTTCAAGCAGTCCCTCCCGAGGCGCTT 266
Db 327 ATTGACAAAATACCAGAGATATAGCAGACTATTCAGCAGTCTGTTCCCGAGGATA 386
Qy 267 CACCTGGAGGCGACACCACTAGGAGGCGGCTTCTGACCGCCACACAGGACAC 326
Db 387 TTCCTGGGAAGAACCACTGACTTTTGAAGACAAGGCAATTTGTCAAAGTGAAGAGTACAT 446

Qy 327 CTCCTTGGAGCGGACTGCTGCTGTACAGGTGAAGTCTTGGGCAACACTTCCCGC 386
Db 447 AAGCATGGAGGAAGACTCTCTTTATCTATGAATTCGTTTATGGATGAACCTTCTCTCC 506
Qy 387 CGACGCGCCGCTGATCGAGCAAGCGCGCTGCGGAGCCCTCCACCGAGATCGTGA 446
Db 507 CAATGGTCCGTTATCGAGAAAACCTTGAAGTGGGAACCATCCATCGAGATTATGA 566
Qy 447 CGAGTGGACGCGCTGCTGCGGCGCAGTCCCTGATGCGCCCTGGAGTCCCGCGGCTCG 506
Db 567 CGTGGCTGATGAGTCTGCTGGAGATATTAGCCATTTCTCTGTTGCTGGAGGAGGTGG 626
Qy 507 CCACCTGACTCCGACCTGACACACCTACCGCTCCAAAGAGCCGCTCCGCGCTCAA 566
Db 627 CAATTACGATGACTTCAAAAGTATTTACAACAAAAA-----GTTGTCAA 677
Qy 567 GATGCCCGCTTCCACTTTCGAGGACCGCATCGAGATCTTCGAGGAGGTGGAGAGG 626
Db 678 ATTGCCAGACTATCACTTTGTGGACCATCGCATTCAGATCTTGAACCATGACAGGATTA 737
Qy 627 CAAGTGTACAGCAGTACGAGCGCGCTGGCGCGCTACT 667
Db 738 CAACAAAGTACGCTGTATGAGATGCTGCTGCTATT 778

RESULT 3

US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

Query Match 22.5%; Score 159.4; DB 4; Length 678;
Best Local Similarity 56.1%; Pred. No. 2.2e-23;
Matches 301; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
Qy 17 TCACCGAGACCATGCTCCCTTCAGGACCACTCGAGGCGACCGTGAACGCGCACTTCA 76
Db 23 TCAAGAGTTTCATAGGTTTAAAGTTTCGATGGAAGAACGTCATCGGCGAGTTG 82
Qy 77 AGTGCAACCGGCAAGCGCGAGGCGCAACCCCTCGAGGGCACCAGGAGATGAAGATCGAGG 136
Db 83 AAATAGAGGCGAAGAGAGAGGAGGCGCATACGAGGCGCACATACCGTAAAGCTTAAG 142
Qy 137 TGATCGAGGCGCGCCCTGCGCTTCGCTTCACATCTCTGTCACCTCTCCATGATGAC 196
Db 143 TAACCAAGGCGGACCTTTCGCTTGGATATTTTGTCCACCAATTTTCAGTATG 202
Qy 197 GCTTCAAGGCGCTTCATCAAGTACGTGCGGCGATCCCCGACTTCTCAAGCAGTCCCTCC 256
Db 203 GAAGCAAGGTATATGTCAGCACCCTCGGACATACAGACTATAAAGCTGTCATTTC 262
Qy 257 CGAGGGCTTCACTCGGAGGCGCACCACTACGAGGAGCGGCGCTTCTGACCGGCC 316

Db 263 CTGAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAAGACGGTGGCGTCGTACTGTAA 322
Qy 317 ACCAGGACACTCCCTGGACGGGACTGCTGTGTACAAAGGTGAAGATCCTGGGCAACA 376
Db 323 CCCAGGATCCAGTTTGCAGGATGCTGTTTCATCTACAAGGTCAAGTTCAITGGGCTGA 382
Qy 377 ACTTCCCGCGAGCGCCCGTGTATCGAGAACAGGCCCGCGCTGGAGCCCTCCACCG 436
Db 383 ACTTTCCTTCGAGTACCTGTTATGCAAAAGAGACAATGGGCTGGGAAGCCAGCACTG 442
Qy 437 AGATCGTGTACGAGGTGGAGCGGCTGTGCGCGCCAGTCCCTGTATGGCCCTGGAGTGCC 496
Db 443 AGGTTTGTATCTCGTGTATGGGCTGTGAAAGAGAGATTATGAAGGCTCTGAAGCTGA 502
Qy 497 CCGCGGTCGCCACCTGACCTGCGCACCTGACACCACTACCGCTCCAAAGAGCCCG 553
Db 503 AAGACGGTGTCTATTACCTAGTTGAATTCAAAAGTATTATACATGGCAAAAGACCTG 559

RESULT 4

US-09-459-956-5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

Query Match 21.9%; Score 154.8; DB 4; Length 699;
Best Local Similarity 54.4%; Pred No. 1.8e-22;
Matches 337; Conservative 0; Mismatches 277; Indels 6; Gaps 1;
Qy 47 TCGAGGCGACCGTGAACGGCCACTTCAAGTGCACCGCAAGGGCGAGGCAACCCCC 106
Db 53 TGAAGGAAACGTTCAATGGGCACTTCTGAAATATAAGGCAAGGAAAGGACAGCCTA 112
Qy 107 TCGAGGCGACCCAGGAGATGAAGATCGAGGTGATCGAGGGGCGCCCTCGCCCTCGCCT 166
Db 113 ATGAAGGCAACCAATACCGTCAACGCTGAGGTATCAAGGGTGGACCTCTCCCAATTTGGTT 172
Qy 167 TCCACATCTCTGCCACCTCTGTCATGACGGCTCCAAAGGCTTCATCAAGTACGTGTCCG 226
Db 173 GGCATATTTTGTGCGCCACAATTCAGTATGGAACAGGCAATTTGCCACCACTGACA 232
Qy 227 GCATCCCGCACTATTCAAGCAGTTCCTCCCGAGGGCTTCACTGGGAGCGGCAACCA 286
Db 233 ACATACATGATTATCTAAAGTGTCAATTCGGAGGGATATACATGGGAACGGTCCATGC 292
Qy 287 CCTACGAGGAGCGGGCTTCTGTACCGCCACACAGGACACCTCCCTGGAGCGGACTGCC 346
Db 293 ACTTTGAAGACGGTGGCTTGTGTGTATCACCATGATATCAGTTTGAAGGCACTGTT 352
Qy 347 TGGTGTACAAGGTGAAGATCTGGGCAACAACTTCCCGCGAGCGGCCCGTGTATGAGA 406
Db 353 TCTACTAGCAGATCAAGTTCACCTGGCTTGAACCTTCTCCAAATGGACCGGTTGTGCAGA 412

Qy 407 ACAAGCGCGCGCTGGAGCCCTCCACGAGATCGTGTACGAGGTGAAGCGGTGCTGC 466
Db 413 AGAAGACAACATGGCTGGGAACCGAGCACTGAGCGTTTGTATCCTCGTATGTTGTGTA 472
Qy 467 CGGCGCAGTCCCTGATGGCCCTGGAGTGCCCGCGGTCGCCACCTGACCTGCCACCTGC 526
Db 473 TAGGAGACATCATCATGCTCTGACAGTTGAAGAGGTGGTCAATACGATGTGACATTA 532
Qy 527 ACACCACCTACGCTCCAGAGCGCGCTCCGCGCTGAGATCGCGGCTTCCACTTCG 586
Db 533 AAATCTGTTTACAGGCGCCAAAGAACCGCCCT-----TGAAGATGCCAGGGTATCACATG 586
Qy 587 AGGACCAACCGCATCGAGATCCTGAGGAGGTGGAGAGGGCAAGTGTCTACAAGCAGTACG 646
Db 587 TTGACACCAACTGGTTATATGGAACAACGACAAAGATTCATGAAGTTGAGGAGCATG 646
Qy 647 AGGCGCGCGTGGCGGCTTAC 666
Db 647 AAATCGCGTTGCAACGCCAC 666

RESULT 5

US-09-459-956-2
; Sequence 2, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Anemonia majano
US-09-459-956-2

Query Match 17.1%; Score 121; DB 4; Length 690;
Best Local Similarity 51.8%; Pred No. 9.3e-16;
Matches 333; Conservative 0; Mismatches 295; Indels 15; Gaps 2;
Qy 22 GAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTCTTCAAGTGC 81
Db 28 GATGACATGAAATGACCTTACCATATGATGGCTGTGTCAATGGGCAATTTACTTTACCGTC 87
Qy 82 ACCGCAAGGGGAGGGCAACCCCTCGAGGGCACCC-----CGAGATGAAGATCGAG 135
Db 88 AAAGGTGAAGGCAACGGGAAGCCATACGAAGGAGCGAGACTTCGACTTTTAAAGTACC 147
Qy 136 GTGATCGAGGGCGGCGCCCTGCGCTTCCGCTTCCACATCTGTCCACCTCTCGCATGTAC 195
Db 148 ATGGCCACGGTGGGCGCCCTTGCATTCCTTTGACATACATCTACAGTGTTCAAATAT 207
Qy 196 GGCCTCAAGCCCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAGAGAGTCCCTC 255
Db 208 GGAATTCGATGTGTTTACTCGGTATCTTACCACTATGCCCGACTATTTCAAACAAGCATTT 267
Qy 256 CCGGAGGGCTTCACTGGGAGCGCACCACTACGAGGAGCGGGCTTCTCTGACCGCC 315
Db 268 CTGACGGAATGATATGAAAGGACTTTTACCTATGAGATGAGAGAGTGTGCTACGCC 327
Qy 316 CACAGGACACCTCCCTCGGACCGGCACTGCTGTTGTATACAGGTGAAGATCTCTGGGCAAC 375

Db 328 AGTGGGAATAGCCTTAAGGCNACTGCTTTAGACCAATCAAGTTTCATGGAGTG 387
 QY 376 AACTTCCCGCGGAGCGCCCGCTGATGAGAACAGCGCCGCTGGAGCCCTCCACC 435
 Db 388 AACTTTCCTGTGTATGAGACCTGTGATGCGAAGAAGAACAACTGGTTGGAGACCACTTTT 447
 QY 436 GAGATCGTGTAGAGTGCAGCGCGTGTGCGCGCCAGTCCCTGATGGCCCTGGAGTGC 495
 Db 448 GAGAAAATGACTGCTCGGATGGAATATTGAAGGTGATGTCACGGCTTCTCATGCTG 507
 QY 496 CCCGGGCTCGCACCTGACCTGCGCACCTGCGCACCACTACCGCTCCAGAAAGCCCGCC 555
 Db 508 CAAGGAGGTGCAATTAACAGATGCCAATTCACACTCTTAAAGACAAAGAACCGG-- 565
 QY 556 TCCGCGCTGAAGATCCCGGCTTTCAGTTCGAGAACACCGCATCGATGCTGGAGGAG 615
 Db 566 -----TGACGATGCCCAAAACCATGTGTGGAACATCGCATTCGAGGACCGACCTT 618
 QY 616 GTGGAGAGGGCAAGTGTACAGCAGTACAGGCGCGCGGTGG 658
 Db 619 GACAAGGTGCAACAGTGTTCAGCTGACGAGGACACGCTGTG 661

RESULT 6

US-08-532-390-40
 ; Sequence 40, Application US/08532390
 ; Patent No. 5795737
 ; GENERAL INFORMATION:
 ; APPLICANT: SEED, BRIAN
 ; APPLICANT: HAAS, JURGEN
 ; TITLE OF INVENTION: High Level Expression of Proteins
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30B
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/532,390
 ; FILING DATE: 22-SEP-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/324,243
 ; FILING DATE: 19-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LECH, KAREN F.
 ; REGISTRATION NUMBER: 35,238
 ; REFERENCE/DOCKET NUMBER: 00786/294001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 542-5070
 ; TELEFAX: (617) 542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 762 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-532-390-40

Query Match 16.3%; Score 115; DB 1; Length 762;
 Best Local Similarity 51.2%; Pred. No. 1.5e-14;
 Matches 352; Conservative 0; Mismatches 320; Indels 15; Gaps 3;
 QY 13 CTGCTGACCGAGACCATGCTCCCTCAGGACCAACCATCGAGGCGCACCGTGAACGGCCACTAC 72

Db 49 CTGTTCAACCGGGTGTGTCCTCATCTCTGTCAGCTGGAGCGGACGTAACGCCACAG 108
 QY 73 TTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCGAGGGCACCCAGAGATGAAGATC 132
 Db 109 TTCAAGTGTCTCGGCGAGGGCGAGGCGATGCACCTACGCGCAAGTCAACCTGAAGTTC 168
 QY 133 GAGGTGATCGAGGGCGGGCCCTGCTGCGCTTCACATCTCCATCTCTCCACCTCTCGCATG 192
 Db 169 ---ATCTGACCAACCGCGCAAGTGTCCGCTGCGCTTGGCCACCTCTGTCACCACTTCAGC 225
 QY 193 TAGGGTCCAAAGGCTTTCATCAAGTACGTGTCCGGCAT-----CCCCGACTACTTCAAG 246
 Db 226 TAGCGGTGCACTGCTTCAGCGCTACCCGACCACTGAAGCAGCAGCACTTCTTCAAG 285
 QY 247 CAGTCCCTCCCGAGGGCTTTCAGCTGGAGGCGCACCACTACGAGGACGGCGGCTTC 306
 Db 286 TCCGCTATGCCGAAAGGTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC 345
 QY 307 CTGACCGCCACAGCAACACCTCCCTGAGCGGCGACTGCTGCTGTACAAGTGAAGATC 366
 Db 346 TACAAGACCGCGCGAGGTGAAGTTCGAGGGCGACACCTGCTGTAACCGCATCGAGCTG 405
 QY 367 CTGGGCAACAACTTCCCGCGCGAGCGCCCGGTGATGTCAGAACAAAGGCGCGCGCTGGGAG 426
 Db 406 AAGGGCATCGACTTCAAGGAGGACGCGCAACATCTCTGGGGCGACAAAGCTGGAGTACAATAC 465
 QY 427 CCTCCACCGAGATCGTGTAGAGGTGAGCGCGCTGCGCGGCGCACTCCCTGATGCGC 486
 Db 466 AACAGCCAAACGTCTATATCATGSCCGACAGCAAGAGAACGGCATCAAGGTGACTTC 525
 QY 487 CTGGAGTCCCGCGCGGTGCGCACCTGACCTGCCACTGCCACACCACTACCGTCCCAAG 546
 Db 526 AAGATCCGGCAC-----AACATCGAGGACGCGAGCTGCGGACCACTACCAAG 579
 QY 547 AAGCCCGCTCCGCTGAAAGTGCCTGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
 Db 580 CAGAACACCCCATCGGCGAGCGGCCCTGCTGTGCTGCCGCAACCACTACTCTGAGCACC 639
 QY 607 CTGGAGGAGGTGGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCTGGCGCGCTAC 666
 Db 640 CAGTCCGCGCTGAGCAAGACCCCAACGAGAGCGCATCATGCTCTCTGCTGAGTTC 699
 QY 667 TCGAGCGCGCGCCCTCCAAAGTGGGC 693
 Db 700 GTGACCGCGCGGATCACTCAAGGC 726

RESULT 7

US-08-717-294-40
 ; Sequence 40, Application US/08717294
 ; Patent No. 6114148
 ; GENERAL INFORMATION:
 ; APPLICANT: SEED, BRIAN
 ; APPLICANT: HAAS, JURGEN
 ; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF
 ; PROTEINS
 ; NUMBER OF SEQUENCES: 110
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Clark & Elbing LLP
 ; STREET: 176 Federal Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA: US/08/717,294
 ; FILING DATE: 20-SEP-1996

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Eibing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/345001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 762 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other
US-08-717-294-40

Query Match 16.3%; Score 115; DB 3; Length 762;
Best Local Similarity 51.2%; Pred. No. 1.5e-14;
Matches 352; Conservative 0; Mismatches 320; Indels 15; Gaps 3;

QY 13 CTGCTGACCGAGACCATGCCCTTCAGGACACCATCGAGGGGACCGTGAACGGCCACTAC 72
Db 49 CTGTTACCGGGGTGTGCCATCCTGGTGAAGTGGACGGGACGTGAACGGCCACAAG 108
QY 73 TTCAAGTGACCGGACCGGAGGCGGAGGCAACCCCTCGAGGGGACCCAGGAGATGAAGATC 132
Db 109 TTCAAGCGTGTCCGCGAGGCGGAGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTC 168
QY 133 GAGGTGATGAGGGGCGGCGCCCTTCGCGCTTCGACATCTGTCGACCTCTGTCATG 192
Db 169 ---ATCTGACACCGGCAAGCTGCCCTGCGCCCTGCGCCACCCCTGCTGACCACTTCAG 225
QY 193 TAGCGTCCAGGCTTCATCAAGTAGTGTCTCGGAT-----CCCGACTACTTCAAG 246
Db 226 TAGCGGTGAGTGCTTACGCGGTACCCCGACACATGAGGAGACGAGTCTTCTCAAG 285
QY 247 CAGTCCCTCCCGAGGCTTCACTGAGGAGCGGACCAACCACTTCAAGGAGCGCGCTTC 306
Db 286 TCCGCCATGCCGAGGCTAGTCCAGGAGCGGACCACTTCTTCAAGGAGCGGCAAC 345
QY 307 CTGACCGCCACAGGACACTCCCTGGAAGGCGGCTGCTGCTGTGTACAGGTGAAGATC 366
Db 346 TACAAGACCGCGCGGAGGTGAAGTTCGAGGGGCGACACCTGTGTGAACCGCATCGAGCTG 405
QY 367 CTGGGCAACAATTCCTCCCGCGGAGGCTTCACCTGGGAGCGGACCACTACGAGAGCGGCTTC 426
Db 406 AAGGGCATCGACTTCAAGGAGGAGCGGCAACATCTTGGGGCACAAGCTGGAGTACAATAC 465
QY 427 CCCTCCACCGAGATCGTGTAGAGGTGGAAGGCGGTGCTGCGGCGGCGGCTCCCTGATGGCC 486
Db 466 AACAGCCACAACGTCTATATCATGCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 525
QY 487 CTGAGGTGCGCGGCGGTTCGCGGCTGACCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 546
Db 526 AAGATCCGCGAC-----AATCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 579
QY 547 AAGCGCGGCTCCGCGCTGAAGATGCGCGGCTTCACTTCGAGGAGGAGGAGGAGGAGGAGGAG 606
Db 580 CAGAACACCCCATCGGCGAGCGGCGGCTGCTGCTGCGCGGAGGAGGAGGAGGAGGAGGAGG 639
QY 607 CTGAGGAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 666
Db 640 CAGTCCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 699
QY 667 TGCAGCGCGCGGCTTCAAGCTGGGC 693
Db 700 GTGACCGCGCGGAGTCACTACGGC 726
```

```

RESULT 8
US-09-486-241-31
; Sequence 31, Application US/09486241
; Patent No. 6472184
; GENERAL INFORMATION:
; APPLICANT: Hegemann, Peter
; TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEIC ACID
; FILE OF INVENTION: POLYMERS
; FILE REFERENCE: 3910/0G706
; CURRENT APPLICATION NUMBER: US/09/486,241
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/EP98/05219
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: DE19736591.4
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified gene from Aquorea victorea
US-09-486-241-31

Query Match 16.2%; Score 114.6; DB 4; Length 717;
Best Local Similarity 52.1%; Pred. No. 1.7e-14;
Matches 358; Conservative 0; Mismatches 314; Indels 15; Gaps 4;

QY 13 CTGCTGACCGAGACCATGCCCTTCAGGACACCATCGAGGGGACCGTGAACGGCCACTAC 72
Db 19 CTGTTACCGGGGTGTGCCATCCTGGTGAAGTGGACGGGACGTGAACGGCCACAAG 78
QY 73 TTCAAGTGACCGGACCGGAGGCGGAGGCAACCCCTCGAGGGGACCCAGGAGATGAAGATC 132
Db 79 TTCTCCGCTCTCCGCGAGGCTGAGGGTGACGCCACCTACGGCAAGCTGACCCCTGAAGTTC 138
QY 133 GAGGTGATGAGGGGCGGCGCCCTGCGCTTCGACATCTTCCACATCTGTCACCTCTGTCATG 192
Db 139 ATCTGCAACAC---CGGCAAGCTGCCCTGCGCCACCCCTGGTCAACCTTACC 195
QY 193 TAGCGTCCAGGCTTCATCAAGTACGTGTCTCGGAT-----CCCGACTACTTCAAG 246
Db 196 TACGGTGTGCTGCTTCTTCCCGCTTACCCCGGACCACTGAAGAGCAGCAGCTTCTTCAAG 255
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGCGGACCACTACGAGAGCGGCGGCTTC 306
Db 256 TCCGCCATGCCGAGGCTTACGTGAGGAGGCGGACCATCTTCTTCAAGGAGCGGCAAC 315
QY 307 CTGACCGCGCCACAGGACACCTTCCCTGGAAGGCGGCTGCTGCTGTGTACAAGTGAAGATC 366
Db 316 TACAAGACCGCGCGGAGGTCAAGTTCGAGGGGCGACACCTGTGTGAACCGCATCGAGCTG 375
QY 367 CTGGGCAACAATTCCTCCCGCGGAGGCTTCACCTGGGAGCGGACCAAGAGGAGGAGGAGG 426
Db 376 AAGGGCATCGACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 432
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGGAAGGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 486
Db 433 TACAACCTCCCAACAGTGTATCATGCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492
QY 487 CTGAGGTGCGCGGCGGCTGCGGACCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 546
Db 493 TTCAAGATCC---GCCACAAACATCGAGGAGCGGCTCCGTGCGAGCTGGCGGACCACTACCAG 549
QY 547 AAGCGCGCTCCGCGCTGAAGATGCCGCTTCCACTTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 606
Db 550 CAGAACACCCCATCGGCGATGGCGCTGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 609
QY 607 CTGAGGAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 666
Db 610 CAGTCCGCGCTGTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 669
```


QY 667 TCCGACGCGCGCCCTTCAAGTGGC 693
 Db 670 GTACCGCTGCGGATCACCACGCG 696

RESULT 9

US-08-818-253-7
 ; Sequence 7, Application US/08818253
 ; Patent No. 5998204
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Miyawaki, Atsushi
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 ; TITLE OF INVENTION: DETECTION OF ANALYTES
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Suite 1400
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: FastSeq for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,253
 ; FILING DATE: 14-MAR-1997
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haile, Ph.D., Lisa A.
 ; REGISTRATION NUMBER: 38,347
 ; REFERENCE/DOCKET NUMBER: 07257/043001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619/678-5070
 ; TELEFAX: 619/678-5099
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1971 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: Coding Sequence
 ; LOCATION: 1....1968

Query Match 16.0%; Score 113; DB 2; Length 1971;

Best Local Similarity 50.9%; Pred. No. 4.1e-14;
 Matches 353; Conservative 0; Mismatches 325; Indels 15; Gaps 3;

QY 13 CTGTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTAC 72
 Db 1279 CTGTTACCGGGGTGGTGCCTCTCTGAGCTGACGGGACGTAACGGCCCAAG 1338
 QY 73 TTCAAGTGCACCGCAAGGCGAGGCAACCCCTCGAGGGCACCGGAGATGAAGATC 132
 Db 1339 TTCAGCTGTCCGCGAGGCGAGGGCGATGCCACCTACGGGAAGTGC 1398
 QY 133 GAGGTGATCGAGGGCGGCCCCCTTCGCTTCCATCTGTCACCTCTCTCATG 192
 Db 1399 ATCTGCACCAAC---CGCAAGCTGCGCGTCCCTTGGCCACCTCTGTGACCCCTGACC 1455
 QY 193 TACGGCTCCAAAGGCGCTTTCATCAAGTAGTGTCCGGCAT-----CCCCGACTACTTCAAG 246
 Db 1456 TACGGGTGAGTGTCTTCAGCGGTACCCCGACCATGAAGCAGCAGCACTTCTTCAAG 1515

QY 247 CAGTCCCTCCCGAGGGCTTCACTCCGAGCGCACCAACCATAGAGACGGCGCTTC 306
 Db 1516 TCCGCAATGCCGAGAGGCTACGTCCAGGAGCGCACATCTTCTTCAAGGACGACGGCAAC 1575
 QY 307 CTGACCGCCACACGAGACACCTCCCTGAGCGGAGTGCCTGGTGTACAAGGTGAAGATC 366
 Db 1576 TACAAGACCCGCGCGAGGTGAAGTTTCGAGGCGGACACCTCGGTGAACCGCATCGAGCTG 1635
 QY 367 CTGGGCAACAATTTCCCGCCGACGCGCCCGTGTGATGAGCAAGCAAGCGCGCTGGGAG 426
 Db 1636 AAGGGCATCGACTTCAAGGAGGACGGCAACATCTCTGGGSCAACACTGAGTACAACTAC 1695
 QY 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGCGTGTCTGCGGGCCAGTCCCTGTATGGCC 486
 Db 1696 AACAGCCACACGTCTATATCATGCGGACAGCAAGAGAACGGCATCAAGGTGAATTC 1755
 QY 487 CTGGAGTCCCGCGCGGTCCGACCTGACCTGCGACCTGCGACACCACTACCGTCCAAAG 546
 Db 1756 AAGATCCGCCAC-----AACATCGAGACGCGAGCTGCGGACCACTTACCAG 1809
 QY 547 AAGCCCGCTCCGCGCTGAAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
 Db 1810 CAGAACACCCCATCGGCGAGCGGCCCGTGTCTGCTGCCGACCACTACTCTGAGCACC 1869
 QY 607 CTGGAGAGGTGGAGAAAGGGCAAGTGTCTACAGCAAGTACGAGGCGCGCGTGGCGCGTAC 666
 Db 1870 CAGTCCGCGCTGAGCAAAAGACCCCAACGAGAGCGCGATCACATGCTCTGTGGAGTTC 1929
 QY 667 TCGGACGCGCGCCCTCCAAAGCTGGGCGCACCAAC 699
 Db 1930 GTGACCGCGCGGGATCACTCTCGGCAAGGAC 1962

RESULT 10

US-08-818-252-7
 ; Sequence 7, Application US/08818252B
 ; Patent No. 6197928
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Miyawaki, Atsushi
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 ; TITLE OF INVENTION: DETECTION OF ANALYTES
 ; FILE REFERENCE: 07257/042001
 ; CURRENT APPLICATION NUMBER: US/08/818,252B
 ; CURRENT FILING DATE: 1997-03-14
 ; NUMBER OF SEQ ID NOS: 56
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1971
 ; TYPE: DNA
 ; ORGANISM: Aequorea victoria
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: {0}...(1968)
 US-08-818-252-7

Query Match 16.0%; Score 113; DB 3; Length 1971;

Best Local Similarity 50.9%; Pred. No. 4.1e-14;
 Matches 353; Conservative 0; Mismatches 325; Indels 15; Gaps 3;

QY 13 CTGTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTAC 72
 Db 1279 CTGTTACCGGGGTGGTGCCTCTCTGAGCTGACGGGACGTAACGGCCCAAG 1338
 QY 73 TTCAAGTGCACCGCAAGGCGAGGCAACCCCTCGAGGGCACCGGAGATGAAGATC 132
 Db 1339 TTCAGCTGTCCGCGAGGCGAGGGCGATGCCACCTACGGGAAGTGC 1398
 QY 133 GAGGTGATCGAGGGCGGCCCCCTTCGCTTCCATCTGTCACATCTCTGTCCACTCTCTCATG 192
 Db 1399 ATCTGCACCAAC---CGCAAGCTGCGCGTCCCTTGGCCACCTCTGTGACCCCTGACC 1455
 QY 193 TACGGCTCCAAAGGCGCTTTCATCAAGTAGTGTCCGGCAT-----CCCCGACTACTTCAAG 246

Db 1456 TACGGCGTGCAGTCTTACGCGGTACCCCGACCATGAGCAGCAGCTCTTCAAG 1515
Qy 247 CAGTCCCTCCCGAGGCTTCACTGCGGAGCGCACACCACTACGAGGACGCGGCTTC 306
Db 1516 TCCGCCATGCGGAAAGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGCAAC 1575
Qy 307 CTGACCGCCACACGAGCACCTCCCTGAGCGGAGCTGCTGCTGATACAGGTGAAGATC 366
Db 1576 TACAGACCGCGCGGAGGTGAAGTTCAGGGGCGACACCTGTGTGAACCCATCGAGCTG 1535
Qy 367 CTGGGCAACAATTTCCCGCGGAGCGGCGGCTGATGAGCAACAAGCGCGCGCTGGGAG 426
Db 1636 AAGGECATCGACTTCAAGGAGGACGCGCAACATCTCGGGGCAAGCTGGAGTCAACTAC 1695
Qy 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCGTGTGCGCGGCCAGTCCCTGATGGCC 486
Db 1696 AACAGCCACACGCTTATATATATATGCTGCGGCAAGCAGAGAACGCGCATCAAGGTGAATTC 1755
Qy 487 CTGAGGTGCGCGCGGCTGCGCACCTGACCTGCGCACCACTACCGCTCCAAG 546
Db 1756 AAGATCGGCAC-----AACATCGAGGACGGCAGCGTGCAGCTCGCGACCACTACCAG 1809
Qy 547 AAGCCCGCTCCGCGCTGAAGATGCGCGGCTTCCACTTTCGAGGACACCGCATCGAGATC 606
Db 1810 CAGAACACCCCATCTCGGCGAGCGGCGCGTGTGCTGCGCGCACCACTACCTGAGCAC 1869
Qy 607 CTGAGGAGGTGGAGAGGCGCAAGTGTACAGCAGTACGAGCGCGCGTGGCGCGCTAC 666
Db 1870 CAGTCCGCTTGAGCAAGACCCCAAGAGAGGCGGATCATGTGCTGCTGGAGTTC 1929
Qy 667 TGGAGCGCGCGCGCTTCCAGTGTGGGCCAAC 699
Db 1930 GTGACCGCGCGGATCACTCTCGGCAAGGAC 1962

RESULT 11

US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: *Ptilosarcus gurneyi*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; FEATURE:
; OTHER INFORMATION: *Ptilosarcus* Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

Query Match 15.9%; Score 112.4; DB 3; Length 1104;
Best Local Similarity 50.8%; Pred. No. 5e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

Qy 16 CTGACCGAGACCATGCGCTTCAGGACCCACCATCGAGGCGACCGTGAACGCCACTACTTC 75
Db 64 CTGAAGAGATTATGTGCGGAAAAGTGTAGCGTTGAGGAAATCGTGAACATCACTCGTTTTT 123

Qy 76 AAGTGCACCGGCAAGCGGCGAGGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
Db 124 TCCATGGAAGGATTTGGAAGGCAATGTATTATTATTGGAACCAATGATGCAATCCGG 183
Qy 136 GTGATCGAGGGGCGGCGCTGCGCTTCGCTTCCACATCCTGTCCACCTCCTGCAATGTAC 195
Db 184 GTTACAAAGGAGGTCCGTTGCCATTCGCTTTCGATATTTGTTCCATAGCTTCCAAATAC 243
Qy 196 GGCTCAAGGCGCTTCATCAAGTACGTTCGCGCATCCCGACTACTTCAAGCAGTCCCTC 255
Db 244 GGGAAATCGCACATTTTACGAAATACCCAGACGACATTTGGGGACTACTTTGTTCAATCATTC 303
Qy 256 CCGGAGGGCTTCACCTGGGAGCGCACCCACCTACGAGGACGGCGGCTTCTGTACCGGCC 315
Db 304 CCGGCTGATTTTCTACGAAGAATCTACGCTTTGAGATGGCGCCATGTTTCACATT 363
Qy 316 CACCAAGGACATCCCTCGAGCGGCGACTGCTGTGTACAAGGTGAAGATCCCTGGGCAAC 375
Db 364 CGTTCAGATATAAGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAAC 423
Qy 376 AACTTCCCGCGACGCGCGCTGTGATGAGACAGGCGCGCGCTGGGAGCCCTCCACC 435
Db 424 GGTTCCTCTAGTAACGAGCCCGCTGATGCAAAAGCCATCCTCGCATGGAGCATCGTTT 483
Qy 436 GAGATCGTGTACGAGGTGGACGCGCTGCTGCGCGCCAGTCCCTGATGCGCCCTGGAGTGC 495
Db 484 GAGGTGGTCTACATCAACAGCGCGCTTCTGTTGGCGGAAGTAGATCTCGTTTACAACTC 543
Qy 496 CCGGCGGTGCGCACCTGACCTGCCACCTGCACACCACTACCGCTCCAA 545
Db 544 GAGTCAGGGAACATATTACTCGTGCACATGAAGACGTTTTTACAGATCCAA 593

RESULT 12

US-09-609-161B-30
; Sequence 30, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: *Ptilosarcus gurneyi*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; OTHER INFORMATION: *Ptilosarcus* Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

Query Match 15.9%; Score 112.4; DB 4; Length 1104;
Best Local Similarity 50.8%; Pred. No. 5e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

Qy 16 CTGACCGAGACCATGCGCTTCAGGACCCACCATCGAGGCGACCGTGAACGCCACTACTTC 75
Db 64 CTGAAGAGATTATGTGCGGAAAAGTGTAGCGTTGAGGAAATCGTGAACATCACTCGTTTTT 123


```
Db 37 CTGAAGAGATTATGTCGCAAAAGCTAGCGTTGAAGGAATCGTGAACAAATCAGCTTTT 96
Qy 76 AAGTGCACCGGCAAGGCGGAGGCGCAACCCCTCGAGGGCAACCCAGGAGATGAAGATCGAG 135
Db 97 TCCATGGAAGGATTTGGAAGAGCAATGATATTATTTGGAAGAACCAATTGATGCAATCCGG 156
Qy 136 GTGATCGAGGGCGGCGCCCTGCGCTTGCCTTCCACATCCTGTCCACCTCCTGCATGTAC 195
Db 157 GTTACAAAGGGAGGTCGCTTGCCTTGCCTTCCGCAATGTTTCCATAGCTTTCCAAATAC 216
Qy 196 GGCTCCAAAGGCCCTTCATCAAGTACGTGTCGGGATCCCGGACTTCAAGCAGTCCCTC 255
Db 217 GGAATCGCACTTTCAGGAATACCCAGACGATTCGGGACTTCTTGTCAATCATTT 276
Qy 256 CCGAGGGCTTACCTGGGAGCGCACCAACCTACGAGACGCGGGCTTCTGACCGCC 315
Db 277 CCGGCTGGATTTTCTACGAAAGAAATCTACGCTTTGAAAGATGGCGCCATTGTTGACATT 336
Qy 316 CACCAGGACACCTCCCTGAGCGGAGCTGCTGCTGTACAAAGTGAAGATCCTGGGCAAC 375
Db 337 CGTTGATATAGTTTAGAGATGATAGTTCCACTACAAAGTGAAGTATAGAGGCAAC 396
Qy 376 AACTTCCCGCCGACGCGCCCGTGTGATGCAAGAACAGGCGGCGCTGGAGCCCTCCAC 435
Db 397 GGTTCCTAGTAACGAGCCGCTGATGCAAAAGCCATCCTCGCATGGAGCCATCGTTT 456
Qy 436 GAGATCGTACGAGTGGAGCGGCTGCTGCGGCGCAGTCCCTGATGCGCCCTGGAGTGC 495
Db 457 GAGGTGCTACATGAAACAGCGCGGCTTCTGGTGGCGAGTAGATCTCGTTTACAAATC 516
Qy 496 CCGGGGCTCGCCACCTGACTGCCACCTGCACACCACTTACCGTCCAA 545
Db 517 GAGTCAGGAACATTACTCGTCCACATGAANAACGTTTACAGATCCAA 566
```

RESULT 15

```
US-09-513-783A-45
; Sequence 45, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 45
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(717)
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-45
```

```
Query Match 15.8%; Score 111.8; DB 4; Length 717;
Best Local Similarity 50.9%; Pred. No. 6.2e-14;
Matches 350; Conservative 0; Mismatches 322; Indels 15; Gaps 3;

Qy 13 CTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTAC 72
Db 22 CTGTTACCGGGGTGTGCCCATCCTGTCGAGTGGACGGCGACGTAAACGGCCCAAG 81
Qy 73 TTCAAGTGCACCGGAGGGGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATC 132
Db 82 TTCAGCGTGTCCGCGGAGGGGAGGCGATGCGACCTACCGGCAAGCTGACCCCTGAAGTTC 141
Qy 133 GAGGTGATGAGGGCGGCGCCCTGCGCTTCCACATCCTGTCCACCTCCTGTCATG 192
```

```
Db 142 ATCTGCACCAC--CGGCAAGCTGCCGTGCCCTGGCCACCCCTCTGTCGACCAACCTGACC 198
Qy 193 TAGGGTCCAAAGGCGCTTCAATCAAGTACGTGTCCGGCAT-----CCCGGACTATTCAAG 246
Db 199 TAGCGCGTGCAGTGTCTTACAGCCGCTACCCCGACACATGAAGCAGCAGCAGCTTCTTCAAG 258
Qy 247 CAGTCCCTCCCGGAGGCTTCACTGCGAGCGCACCACTACCTACGAGGACGGGGCTTC 306
Db 259 TCCGGCATGCCCGAAGGCTAGCTCAGAGGCGCACCATCTTCTTCAAGGACGACGGCAAC 318
Qy 307 CTGACCGCCCAACAGGACACCTCCTGGAACGCGGACTGCTGCTGTGTTACAAGGTGAAGATC 366
Db 319 TACAAGACCGCGCGGAGTGAAGTTCGAGGGCGACACCTCTGTGTGAACCGCATCGAGCTG 378
Qy 367 CTGGGCAACAATTTCCCGCGAGCGGCCCTGTGATGAGAACAGGCAAGGCGCGCTGGGAG 426
Db 379 AAGGCGATCGACTTCAAGGAGGACGGCAACATCTCTGGGCGACAAGCTGGAGTACAATAC 438
Qy 427 CCCTCCACCGGAGATCGTGTACGAGGTGAGCGGCGTGTGCGCGGCCCAAGTCCCTGATGCC 486
Db 439 AACAGCCACAACGTCTATATCATGGCCGACAAAGCAGAAGAACGGCATCAAGGTGAACCTC 498
Qy 487 CTGAGTGCCTCCGGGGGTGCGCACCTGACCTGCCACCTGCAACACCACTACCGCTCCAAG 546
Db 499 AAGATCCGCCAC-----AACATCGAGACGCGAGCTGCGCGCACTACTACCG 552
Qy 547 AAGCCGCGCTCCGCGCTGAGATGCCCGGCTTCCACTTTCGAGGACCAACCGCATCGAGATC 606
Db 553 CAGAACACCCCATCGGCGACGGCCCGTGTCTGCTGCCCGACACCACTACTCTGAGCACC 612
Qy 607 CTGAGGAGGTGGAGAGGCGCAAGTGTCTAACGAGATGACGAGGCGCGCTGGCGGCTAC 666
Db 613 CAGTCCGCGCTGAGCAAAAGACCCCAACGAGAAGCGCGATCATGTGTCCTCTCTGAGTTC 672
Qy 667 TCGGACGCGCGCCCTCCCAAGCTGGGC 693
Db 673 GTGACCGCGCGGGATCCTCTCGGC 699
```

Search completed: August 3, 2004, 15:50:33
Job time : 73.4657 secs

Blank Sheet USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 14:04:25 ; Search time 514.043 Seconds
(without alignments)
6743.643 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 707.

Sequence: 1 ggatcgctcctcgtgac.....ctggccacaactgaagctt 707

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	707	100.0	707	14	US-10-006-922-39
2	707	100.0	707	15	US-10-081-864-21
3	678.4	96.0	699	14	US-10-006-922-41
4	669	94.6	696	17	US-10-724-178-17
5	649.2	91.8	634	15	US-10-081-864-23
6	571.4	80.8	591	17	US-10-724-178-1046
7	532.6	75.3	555	17	US-10-724-178-1064
8	481.6	68.1	504	17	US-10-724-178-1060
9	440.2	62.3	696	14	US-10-006-922-13
10	440.2	62.3	696	15	US-10-081-864-9
11	438.4	62.0	456	17	US-10-724-178-1052
12	436.8	61.8	767	15	US-10-155-809-1
13	435.2	61.6	767	15	US-10-155-809-3
14	433.6	61.3	767	15	US-10-155-809-5

15	417.8	59.1	687	9	US-09-976-673-11	Sequence 11, Appl
16	417.8	59.1	687	9	US-09-976-673-13	Sequence 13, Appl
17	417.8	59.1	1396	9	US-09-976-673-15	Sequence 15, Appl
18	417.8	59.1	1424	9	US-09-976-673-17	Sequence 17, Appl
19	387.6	54.8	405	17	US-10-724-178-1050	Sequence 1050, Ap
20	345.6	48.9	363	17	US-10-724-178-1058	Sequence 1058, Ap
21	323.4	45.7	333	17	US-10-724-178-1056	Sequence 1056, Ap
22	319.6	45.2	649	17	US-10-439-262-12	Sequence 12, Appl
23	315	44.6	675	14	US-10-066-922-38	Sequence 38, Appl
24	315	44.6	675	15	US-10-081-864-13	Sequence 13, Appl
25	315	44.6	1050	14	US-10-060-857-7	Sequence 7, Appl
26	313.4	44.3	678	14	US-10-006-922-36	Sequence 36, Appl
27	313.4	44.3	678	15	US-10-081-864-14	Sequence 14, Appl
28	313.4	44.3	678	15	US-10-315-920-5	Sequence 5, Appl
29	313.4	44.3	681	15	US-10-121-258-7	Sequence 7, Appl
30	312.6	44.2	675	17	US-10-724-178-15	Sequence 15, Appl
31	312.6	44.2	678	15	US-10-121-258-9	Sequence 9, Appl
32	310.2	43.9	678	15	US-10-081-864-7	Sequence 7, Appl
33	310.2	43.9	678	15	US-10-121-258-5	Sequence 5, Appl
34	310.2	43.9	678	15	US-10-315-920-1	Sequence 1, Appl
35	308.6	43.6	678	15	US-10-121-258-3	Sequence 3, Appl
36	308.6	43.6	681	15	US-10-121-258-3	Sequence 3, Appl
37	308.6	43.6	681	15	US-10-121-258-23	Sequence 23, Appl
38	308.6	43.6	1638	15	US-10-214-932-51	Sequence 51, Appl
39	308.6	43.6	1647	15	US-10-214-932-75	Sequence 75, Appl
40	308.6	43.6	4692	15	US-10-161-403-29	Sequence 29, Appl
41	308.6	43.6	4692	17	US-10-433-640-16	Sequence 16, Appl
42	308.6	43.6	6984	14	US-10-001-189-45	Sequence 45, Appl
43	307	43.4	681	14	US-10-006-922-35	Sequence 35, Appl
44	307	43.4	681	14	US-10-006-922-37	Sequence 37, Appl
45	307	43.4	723	15	US-10-152-296-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-006-922-39
Sequence 39, Application US/10006922
Publication No. US20020197676A1
GENERAL INFORMATION:
APPLICANT: Lukyanov, Sergey A
APPLICANT: Fradkov, Arcady F.
APPLICANT: Labas, Yulii A.
APPLICANT: Matz, Mikhail V.
APPLICANT: Tersikh, Alexey
TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
FILE REFERENCE: CLON-035CIP
CURRENT APPLICATION NUMBER: US/10/006,922
CURRENT FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 09/120,330
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/457,898
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,144
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,477
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/457,556
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/444,338
PRIOR FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 707
TYPE: DNA
ORGANISM: Anemonia sulcata
US-10-006-922-39

Query Match 100.0%; Score 707; DB 14; Length 707;
Best Local Similarity 100.0%; Pred. No. 1e-159;

Matches 707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCCGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCAACCATTCGAGGGCACCGTG 60
DB 1 GGATCCGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCAACCATTCGAGGGCACCGTG 60

QY 61 AACGGCCACTACTTCAAGTGTGACCGGCAAGGGCGAGGGCAACCCCTTCGAGGGCACCCAG 120
DB 61 AACGGCCACTACTTCAAGTGTGACCGGCAAGGGCGAGGGCAACCCCTTCGAGGGCACCCAG 120

QY 121 GAGATGAAGTTCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180
DB 121 GAGATGAAGTTCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

QY 181 ACCTCTGCTGATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTAC 240
DB 181 ACCTCTGCTGATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTAC 240

QY 241 TTCAGGAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCCACCTACGAGGACGGC 300
DB 241 TTCAGGAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCCACCTACGAGGACGGC 300

QY 301 GGCTTCTGACCGGCCACAGGACACCTCCCTGAGCGGGCGACTGCTGTGTACAAAGTG 360
DB 301 GGCTTCTGACCGGCCACAGGACACCTCCCTGAGCGGGCGACTGCTGTGTACAAAGTG 360

QY 361 AAGATCTTGGGCAACACTTCCCGCGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 420
DB 361 AAGATCTTGGGCAACACTTCCCGCGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 420

QY 421 TGGGAGCGCTCCACCGAGATCGTGTACGAGGTGACGGCGGTGCTGCGCGGCGAGTCCCTG 480
DB 421 TGGGAGCGCTCCACCGAGATCGTGTACGAGGTGACGGCGGTGCTGCGCGGCGAGTCCCTG 480

QY 481 ATGGCCCTGAGTGCCTGCGGGGTGCGGACCTGACCTGCGACCTGCGACACCTACCGC 540
DB 481 ATGGCCCTGAGTGCCTGCGGGGTGCGGACCTGACCTGCGACCTGCGACACCTACCGC 540

QY 541 TCCAGAGCGCGCTCCGCGCTGAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATC 600
DB 541 TCCAGAGCGCGCTCCGCGCTGAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATC 600

QY 601 GAGATCTTGGAGGAGTGGAGAGGCAAGTGTCTACAGCAGTACGAGGCGCGGTGGGC 660
DB 601 GAGATCTTGGAGGAGTGGAGAGGCAAGTGTCTACAGCAGTACGAGGCGCGGTGGGC 660

QY 661 CGCTACTGCGAGCGCGCGCGCTCCAAAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 707
DB 661 CGCTACTGCGAGCGCGCGCGCTCCAAAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 707

RESULT 2

US-10-081-864-21
; Sequence 21, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexander
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/005,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21

QY 1 GGATCCGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCAACCATTCGAGGGCACCGTG 60
DB 1 GGATCCGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCAACCATTCGAGGGCACCGTG 60

QY 61 AACGGCCACTACTTCAAGTGTGACCGGCAAGGGCGAGGGCAACCCCTTCGAGGGCACCCAG 120
DB 61 AACGGCCACTACTTCAAGTGTGACCGGCAAGGGCGAGGGCAACCCCTTCGAGGGCACCCAG 120

QY 121 GAGATGAAGTTCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180
DB 121 GAGATGAAGTTCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

QY 181 ACCTCTGCTGATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTAC 240
DB 181 ACCTCTGCTGATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTAC 240

QY 241 TTCAGGAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCCACCTACGAGGACGGC 300
DB 241 TTCAGGAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCCACCTACGAGGACGGC 300

QY 301 GGCTTCTGACCGGCCACAGGACACCTCCCTGAGCGGGCGACTGCTGTGTACAAAGTG 360
DB 301 GGCTTCTGACCGGCCACAGGACACCTCCCTGAGCGGGCGACTGCTGTGTACAAAGTG 360

QY 361 AAGATCTTGGGCAACACTTCCCGCGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 420
DB 361 AAGATCTTGGGCAACACTTCCCGCGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 420

QY 421 TGGGAGCGCTCCACCGAGATCGTGTACGAGGTGACGGCGGTGCTGCGCGGCGAGTCCCTG 480
DB 421 TGGGAGCGCTCCACCGAGATCGTGTACGAGGTGACGGCGGTGCTGCGCGGCGAGTCCCTG 480

QY 481 ATGGCCCTGAGTGCCTGCGGGGTGCGGACCTGACCTGCGACCTGCGACACCTACCGC 540
DB 481 ATGGCCCTGAGTGCCTGCGGGGTGCGGACCTGACCTGCGACCTGCGACACCTACCGC 540

QY 541 TCCAGAGCGCGCTCCGCGCTGAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATC 600
DB 541 TCCAGAGCGCGCTCCGCGCTGAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATC 600

QY 601 GAGATCTTGGAGGAGTGGAGAGGCAAGTGTCTACAGCAGTACGAGGCGCGGTGGGC 660
DB 601 GAGATCTTGGAGGAGTGGAGAGGCAAGTGTCTACAGCAGTACGAGGCGCGGTGGGC 660

QY 661 CGCTACTGCGAGCGCGCGCGCTCCAAAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 707
DB 661 CGCTACTGCGAGCGCGCGCGCTCCAAAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 707

RESULT 3

US-10-006-922-41
; Sequence 41, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
; FILE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP

;; CURRENT APPLICATION NUMBER: US/10/006,922
;; CURRENT FILING DATE: 2001-12-04
;; PRIOR APPLICATION NUMBER: 09/120,330
;; PRIOR FILING DATE: 1998-12-11
;; PRIOR APPLICATION NUMBER: 09/457,898
;; PRIOR FILING DATE: 1999-12-09
;; PRIOR APPLICATION NUMBER: 09/458,144
;; PRIOR FILING DATE: 1999-12-09
;; PRIOR APPLICATION NUMBER: 09/458,477
;; PRIOR FILING DATE: 1999-12-09
;; PRIOR APPLICATION NUMBER: 09/457,556
;; PRIOR FILING DATE: 1999-12-09
;; PRIOR APPLICATION NUMBER: 09/444,338
;; PRIOR FILING DATE: 1999-11-19
;; NUMBER OF SEQ ID NOS: 46
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 41
;; LENGTH: 699
;; TYPE: DNA
;; ORGANISM: Anemonia sulcata
US-10-006-922-41

Query Match 96.0%; Score 678.4; DB 14; Length 699;
Best Local Similarity 98.4%; Pred. No. 7.2e-153;
Matches 685; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGACCGGTGAACGGC 66
Db |||||
QY 4 GCCTCCCTGCTGAAGAGACCATGCCCTTCAAGACCAACCATCGAGGACCGGTGAACGGC 63
Db |||||
QY 67 CACTACTTCAGTGCACCGGACGAGGCGGCAACCCCTCGAGGACCGGAGGATG 126
Db |||||
QY 64 CACTACTTCAGTGCACCGGACGAGGCGGCAACCCCTTCAGGACCGGAGGATG 123
Db |||||
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTGCCTTCACATCTCTGTCACCTCC 186
Db |||||
QY 124 AAGATCGAGGTGATCGAGGCGGCGCCCTTGCCTTCACATCTCTGTCACCTCC 183
Db |||||
QY 187 TGCATGTACGGCTCCAGAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db |||||
QY 184 TGCATGTACGGCTCCAGAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 243
Db |||||
QY 247 CAGTCCCTCCCGAGGCGGCTTCACTGGGAGCGCACCAACCTACGAGGACCGGCGCTTC 306
Db |||||
QY 244 CAGTCCCTCCCGAGGCGGCTTCACTGGGAGCGCACCAACCTACGAGGACCGGCGCTTC 303
Db |||||
QY 307 CTGACCGGCGGACCGGACCTCCCTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 366
Db |||||
QY 304 CTGACCGGCGGACCGGACCTCCCTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 363
Db |||||
QY 367 CTGGGCAACAACTTCCCGCGGCGGCTGATGATGAGAAACGAGGCGGCGCTGGGAG 426
Db |||||
QY 364 CTGGGCAACAACTTCCCGCGGCGGCTGATGATGAGAAACGAGGCGGCGCTGGGAG 423
Db |||||
QY 427 CCTTCCACCGAGATCGTGTACGAGGTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 486
Db |||||
QY 424 CCTTCCACCGAGATCGTGTACGAGGTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 483
Db |||||
QY 487 CTGGAGTCCCGCGGCTGCGACCTGACCTGCCACCTGACACCACTACCGCTCCAAAG 546
Db |||||
QY 484 CTGGAGTCCCGCGGCTGCGACCTGACCTGCCACCTGACACCACTACCGCTCCAAAG 543
Db |||||
QY 547 AAGCCCGGCTCCCGCTGAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db |||||
QY 544 AAGCCCGGCTCCCGCTGAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 603
Db |||||
QY 607 CTGGAGGAGGTGGAAGAGGCGGAGTGTACAGGAGTACGAGGCGGCGCTGGGCGCTAC 666
Db |||||
QY 604 ATGGAGGAGGTGGAAGAGGCGGAGTGTACAGGAGTACGAGGCGGCGCTGGGCGCTAC 663
Db |||||
QY 667 TCGAAGCGCGGCGGCTCCAAAGTCCGGGCAACAATGA 702
Db |||||
QY 664 TCGAAGCGCGGCGGCTCCAAAGTCCGGGCAACAATGA 699
Db |||||

RESULT 4

US-10-724-178-17
;; Sequence 17, Application US/10724178
;; Publication No. US20040137528A1
;; GENERAL INFORMATION:
;; APPLICANT: Odysey Thera, Inc.
;; APPLICANT: Michnick, Stephen
;; APPLICANT: Macdonald, Marnie
;; APPLICANT: Lamerdin, Jane
;; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
;; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
;; FILE REFERENCE: ODDY007
;; CURRENT APPLICATION NUMBER: US/10/724,178
;; CURRENT FILING DATE: 2003-12-01
;; PRIOR APPLICATION NUMBER: US 60/461,133
;; PRIOR FILING DATE: 2003-04-09
;; NUMBER OF SEQ ID NOS: 1067
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 17
;; LENGTH: 696
;; TYPE: DNA
;; ORGANISM: Artificial
;; FEATURE:
;; OTHER INFORMATION: "KFPI" a mutated variant of kindling fluorescent protein asCP fro
;; OTHER INFORMATION: m A. sulcata
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(696)
US-10-724-178-17
Query Match 94.6%; Score 669; DB 17; Length 696;
Best Local Similarity 97.8%; Pred. No. 1.3e-150;
Matches 678; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGACCGGTGAACGGC 66
Db |||||
QY 4 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGACCGGTGAACGGC 63
Db |||||
QY 67 CACTACTTCAGTGCACCGGACGAGGCGGCAACCCCTCGAGGACCGGAGGATG 126
Db |||||
QY 64 CACTGCTTCAAGTGCATCGGCAAGGCGGAGGCAACCCCTTCGAGGACCGGAGGATG 123
Db |||||
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTGCCTTCACATCTCTGTCACCTCC 186
Db |||||
QY 124 AAGATCGAGGTGATCGAGGCGGCGCCCTTGCCTTCACATCTCTGTCACCTCC 183
Db |||||
QY 197 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db |||||
QY 184 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 243
Db |||||
QY 247 CAGTCCCTCCCGAGGCGGCTTCACTGGGAGCGCACCAACCTACGAGGACCGGCGCTTC 306
Db |||||
QY 244 CAGTCCCTCCCGAGGCGGCTTCACTGGGAGCGCACCAACCTACGAGGACCGGCGCTTC 303
Db |||||
QY 307 CTGACCGGCGGACCGGACCTCCCTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 366
Db |||||
QY 304 CTGACCGGCGGACCGGACCTCCCTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 363
Db |||||
QY 367 CTGGGCAACAACTTCCCGCGGCGGCTGATGATGAGAAACGAGGCGGCGCTGGGAG 426
Db |||||
QY 364 CTGGGCAACAACTTCCCGCGGCGGCTGATGATGAGAAACGAGGCGGCGCTGGGAG 423
Db |||||
QY 427 CCTTCCACCGAGATCGTGTACGAGGTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 486
Db |||||
QY 424 CCTTCCACCGAGATCGTGTACGAGGTGGAAGGAGTGCCTGTGTACAGGTGAAGATC 483
Db |||||
QY 487 CTGGAGTCCCGCGGCTGCGACCTGACCTGCCACCTGACACCACTACCGCTCCAAAG 546
Db |||||
QY 484 CTGGAGTCCCGCGGCTGCGACCTGACCTGCCACCTGACACCACTACCGCTCCAAAG 543
Db |||||
QY 547 AAGCCCGGCTCCCGCTGAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db |||||

Db 303 GGTGGCGGCTGGAGCCCGAAACCGAGATCGTGTACAGGTGGAGCGGCTGCTGGCGG 362
Qy 471 CCAGTCCTCATGGCTGAGTGGCCCGCGGTGCCACCTGACCTGACCTGACCTGACAC 530
Db 363 CCAGTCCTCATGGCTGAGTGGCCCGCGGTGCCACCTGACCTGACCTGACAC 422
Qy 531 CACCTACCGCTCCAAGAAGCCCGCTCCGCGCTGAAAGATGCGCGGTTCACATTCGAGGA 590
Db 423 CACCTACCGCTCCAAGAAGCCCGCTCCGCGCTGAAAGATGCGCGGTTCACATTCGAGGA 482
Qy 591 CCACCGCATGAGATCTGAGAGAGTGGAGAGGCGAGTGTACAAAGCATGACAGGC 650
Db 483 CCACCGCATGAGATCTGAGAGAGTGGAGAGGCGAGTGTACAAAGCATGACAGGC 542
Qy 651 CGCGTGGGCGGTACTGACAGCGCGCCCGCTCCAAAGCTGGGCGCAAC 699
Db 543 CGCGTGGGCGGTACTGACAGCGCGCCCGCTCCAAAGCTGGGCGCAAC 591

RESULT 7

US-10-724-178-1064
; Sequence 1064, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1064
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F1F, with position 1 Met removed
; NAME/KEY: CDS
; LOCATION: (1)..(555)
; OTHER INFORMATION: KFP F1F corresponds to aa residues 1-186 of KPFI
US-10-724-178-1064

Query Match 75.3%; Score 532.6; DB 17; Length 555;
Best Local Similarity 97.5%; Pred. No. 5.2e-118;
Matches 541; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Qy 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACATCGAGGCGACCGTGAACGGC 66
Db 1 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACATCGAGGCGACCGTGAACGGC 60
Qy 67 CACTACTTCAAGTGCACCGGAGGGGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db 61 CACTGTCTCAAGTGCATCGGCAAGGGCGAGGCAACCCCTTCAGGCGACCCAGGAGATG 120
Qy 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCACATCTTCCACATCTTCCACCTCC 186
Db 121 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCACATCTTCCACATCTTCCACCTCC 180
Qy 187 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 181 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 240
Qy 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGCACCAACCTACGAGGACGCGGCTTC 306
Db 241 CAGTCCCTCCCGAGGCTTCACCTGGAGCGCACCAACCTACGAGGACGCGGCTTC 300

Qy 307 CTGACCGCCCAACAGGACACCTCCCTGGACGCGGACCTGCTGTGTACAAGTGAAGATC 366
Db 301 CTGACCGCCCAACAGGACACCTCCCTGGACGCGGACCTGCTGTGTACAAGTGAAGATC 360
Qy 367 CTGGCAACAATCTCCCGCGGACGCGCCCTGATGAGAAACAAGGCGCGCTGGAG 426
Db 361 CTGGGTAAACAATCTCCCGCGGACGCGCCCTGATGAGAAACAAGGCGCGCTGGAG 420
Qy 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGCTGCGGCGCAGTCCCTGATGGCC 486
Db 421 CCGGAAACCGAGATCGTGTACGAGGTGAGCGGCTGCTGCGGCGCAGTCCCTGATGGCC 480
Qy 487 CTGAGTGCCTCCCGCGGCTGCGCACTGACCTGCGACCTGCAACCACTACCGCTCCAAG 546
Db 481 CTGAAGTGCCTCCCGCGGCTGCGCACTGACCTGCGACCTGCAACCACTACCGCTCCAAG 540
Qy 547 AAGCCCGCTCGGC 561
Db 541 AAGCCCGCTCGGC 555

RESULT 8

US-10-724-178-1060
; Sequence 1060, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1060
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F1E, with position 1 Met removed
; NAME/KEY: CDS
; LOCATION: (1)..(504)
; OTHER INFORMATION: KFP F1E corresponds to aa residues 1-169 of KPFI
US-10-724-178-1060

Query Match 68.1%; Score 481.6; DB 17; Length 504;
Best Local Similarity 97.2%; Pred. No. 8.1e-106;
Matches 490; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Qy 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACATCGAGGCGACCGTGAACGGC 66
Db 1 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACATCGAGGCGACCGTGAACGGC 60
Qy 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db 61 CACTGTCTCAAGTGCATCGGCAAGGGCGAGGCAACCCCTTCGAGGGCGACCCAGGAGATG 120
Qy 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCACATCTTCCACATCTTCCACCTCC 186
Db 121 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCACATCTTCCACATCTTCCACCTCC 180
Qy 187 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 181 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 240
Qy 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGCACCACTACGAGGACGCGGCTTC 306

241	DB	CAGTCTTCCCGAGGGCTTACCTGGAGCGCACACCACTTACGAGNACGCGCGCTTC	300
307	QY	CTGACGCCCAACAGGACACCTCCCTGGACGGGACTGCTGTGTGTAACAGGTGAAGATC	366
301	DB	CTGACGCCCAACAGGACACCTCCCTGGACGGGACTGCTGTGTGTAACAGGTGAAGATC	360
367	QY	CTGGGCAACAACATTCTCCCGCGGACGGGCCCGGTGATGAGAACAGGCCCGCCCTGGGAG	426
361	DB	CTGGGTAAACAATTCTCCCGCGGACGGGCCCGGTGATGAGAACAGGTGCGCGCTGGGAG	420
427	QY	CCCTCCACGAGATCGTGTACGAGGTGGAACGGCGCTGTGCGCGGCACAGTCCCTGATGGCC	486
421	DB	CCCGGNAACGAGATCGTGTACGAGGTGGAACGGCGCTGTGCGCGGCACAGTCCCTGATGGCC	480
487	QY	CTGGAGTGCCTCCGGCGGTGCGCAC	510
481	DB	CTGAAGTGCCCGCGGCGCGCAC	504

RESULT 9

```

US-10-006-922-13
; Sequence 13, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Pradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922

```

```

; ORGANISM: Anemomilla sulcata
US-10-006-922-13

Query Match      62.3%;      Score 440.2;      DB 14;      Length 695;
Best Local Similarity 77.2%;      Pred. No. 6.4e-96;
Matches 535;      Conservative 0;      Mismatches 158;      Indels 0;      Gaps 0;

```

Qy	67	C	A	C	T	A	C	T	T	C	A	G	T	G	C	A	C	G	G	C	G	G	C	A	C	C	C	C	T	C	G	A	G	G	C	A	C	C	C	A	G	A	G	A	G	T	126				
Db	64	C	A	C	T	A	C	T	T	C	A	G	T	G	T	A	C	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	123						
Qy	127	A	A	G	A	T	C	G	A	G	G	T	G	A	T	C	G	A	G	G	G	G	C	C	C	C	C	C	T	T	G	C	C	T	T	C	A	C	A	T	C	T	G	C	C	A	C	C	T	186	
Db	124	A	A	G	A	T	A	G	G	T	C	A	T	C	G	A	A	G	A	G	G	T	C	C	A	T	T	G	C	C	A	T	T	G	C	C	A	T	T	T	T	G	T	C	A	C	A	G	A	T	183
Qy	187	T	C	A	T	G	T	A	C	G	G	T	C	C	A	A	G	C	C	T	T	C	A	T	C	A	G	T	G	T	C	C	G	G	A	T	C	C	C	G	A	C	T	A	C	T	T	C	A	G	246
Db	184	T	C	T	A	T	G	T	A	C	G	T	A	G	A	A	G	C	C	T	T	C	A	T	C	A	A	G	T	A	T	G	T	G	T	C	A	G	A	A	T	C	T	C	G	A	C	T	A	C	243

247	Qy	CHGTCCCTCCCGAGGGCTTCACTCGGAGCGCACCAACCTACGAGAGCGCGGCTTC	303
248	Qy		
249	Qy		
244	Db	CHGCTTTTCCCTGAAGSTTTTACTTTGGGAAGAACCACAACTTACGAGGATGGAGGCTTT	303
245	Db		
307	Qy	CTGACGCGCCACACAGGACACCTCCCTGGACGGCGACTGCTGTGTATCAAGGTGAAGATC	366
308	Qy		
304	Db	CTTACAGCTCATCAGGACAAAGCTTAGTGGAGATTGCTCGTTTACAAGGTCAAGATT	363
309	Db		
367	Qy	CTGGGCACAACATTTCCCGCGGACGGCCCGTGATGCGAGAAACAGGCCCGCGCTGGGAG	426
368	Qy		
364	Db	CTTGGTAATAATTTTCTCGTGTATGGGCCCGCTGATGAGAAACAAAGCAGAGATGGGAG	423
369	Db		
427	Qy	CCCTCCACCGAGATCGTGTAAGGTGGAGCGGCTGCTCGGGCGGACAGTCCCTGATGCC	486
428	Qy		
424	Db	CATCCACCGAGATAGTTTATGAAGTTGACGGTGTCTCTGTGTGAACAGTCTTTGATGCC	483
429	Db		
487	Qy	CTGGAGTGCCTCCGGCGGTTCGCCACTGACCTGCCACCTGCACACCACTTACCGCTCCAAG	546
488	Qy		
484	Db	CTTAGTGCCTCGTGTGCTGCTATCTGACTTGCATCTCCATACTACTTACAGGTCCAAA	543
489	Db		
547	Qy	AAGCCCGCTCCGGCTGTAAAGATGCCCGGCTTCCACTTTGAGGACCAACCGCATCCAGATC	606
548	Qy		
544	Db	AAACAGCTGCTGCCTTTGAAGATGCCAGGATTTTCATTTTGAAGATCATCCCATCCAGATA	603
549	Db		
607	Qy	CTGGAGGAGGTGGAGAGGGGCAAGTGCTACAGCAGGTACGAGCGCCCGCTGGCGCGCTAC	666
608	Qy		
604	Db	ATGGAGGAGTTTGAAAGGCAAGTGTATTAACAGTACGAAGCAGCAGTGGGCGAGGTAC	663
609	Db		
667	Qy	TGCGACGCCGCCCTCCAAAGCTGGGCCAAC	699
668	Qy		
664	Db	TGTGATGCTGCTCCACTCAAGCTTGGACATAAC	696
665	Db		

RESIT.T 10

```

US-10-081-864-9
; Sequence 9, Application US/10081864
; Publication NO. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexander
; APPLICANT: Pradtov, Arcady
; TITLE OF INVENTION: NO. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Anemonia sulcata
US-10-081-864-9

```

	Query Match	62.3%	Score 440.2	DB 15	Length 696	
	Best Local Similarity	77.2%	Pred. No. 6.4e-36			
	Matches 535	Conservative 0	Mismatches 158	Indels 0	Gaps 0	
QY	7	GCCTCCCTGTGTCACCGAGACCATGCGCTTCAGGACACCATCGAGGGACCGTGAACGCG	66			
Db	4	GCTTCCTCTTTTAAAGAGAAGCTATGCGCTTTTAAAGACACCATTTGAAGGGACGGTTTAAATGCG	63			
QY	67	CACCTACTTTCAGGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGGACCCACAGAGATG	126			
Db	64	CACCTACTTTCAGGTGCACGAAAGAGAGGGCAACCCATTTGAGGTTCAGCAGGAAATG	123			

```
QY 127 AAGATCGAGGTGATCGAGGCGCGCCCTCGCCCTTCCACATCTGTCACCTCC 186
    |||||
Db 124 AAGATAGAGTGCATCGAAGGAGTCCATTGGCCATTGCCATTTTGTCAAGAGT 183
    |||||
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTAGTGGTCCGGCATGCCGACTACTTCAAG 246
    |||||
Db 184 TGTATGTACGGTAGTAAGGCTTCATCAAGTAGTGTGTAGGAAATTCCTGACTACTTCAAG 243
    |||||
QY 247 CAGTCCCTCCCGAGGGCTTCACTGGGAGCGCCACCACTTACGAGGACGGCGGCTTC 306
    |||||
Db 244 CAGTCTTTCCCTGAAGGTTTACTTGGGAAGAACCACTTACGAGGATGGAGCTTT 303
    |||||
QY 307 CTGACCGCCACACAGACACTCCCTGGACGGGACTGCTGGTGTACAAAGTGAAGATC 366
    |||||
Db 304 CTTACAGCTCATCAGGACACAAGCTTAGATGGAGATTGCTGCTTTTACAAAGTCAAGATT 363
    |||||
QY 367 CTGGGCAACAATTTCCCGCCGAGCGCCGCTGATGACAGAACAAAGCGCGCGCTGGAG 426
    |||||
Db 364 CTTGGTAATTAATTTCTGTATGGCCCGCTGATGACAGAACAAAGCAGAGATGGAG 423
    |||||
QY 427 CCTCCACCGAGATCGTGTACGAGGTGGACGGCGTCTCGCGGCGAGTCCCTGATGGCC 486
    |||||
Db 424 CCATCCACCGAGATGTTTATGAAGTTGACGGTGTCTCGTGGACAGTCTTTGATGGCC 483
    |||||
QY 487 CTGGAGTGGCCCGCGGTGCGCACTGACCTGCCACTGCACACCACTACCGCTCCAG 546
    |||||
Db 484 CTTAAGTGGCCCTGGTGGTCTGCTACTGACTTGCCTATCTTCACTACTTACAGGTCCTAAA 543
    |||||
QY 547 AAGCCGCGCTCCCGCTCGAAGATGCGCGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
    |||||
Db 544 AACCAGCTGCTGCTTGAAGATGCCAGGATTTTCAATTTGAAGATCATCGCATCGAGATA 603
    |||||
QY 607 CTGGAGAGGTGAGAAAGGCAAGTGTCAAGCAGTACGAGCGCGCTGGCGCGCTAC 666
    |||||
Db 604 ATGGAGGAAGTTGAGAAAGCAAGTGTCTATAACAGTACGAGCAGCAGTGGCGCAGGTAC 663
    |||||
QY 667 TGGAGCGCGCCCTCCAGCTGGGCGCACAC 699
    |||||
Db 664 TGTGATGCTGCTCCATCCAGCTTGACATAC 696
    |||||
```

RESULT 11

```
US-10-724-178-1052
; Sequence 1052, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odysey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: MacDonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: COMPLEMENTATION ASSAYS
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1052
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F1C, with position 1 Met removed
; NAME/KEY: CDS
; LOCATION: (1)..(456)
; OTHER INFORMATION: KFP F1C corresponds to aa residues 1-153 of KPFI
US-10-724-178-1052
```

Query Match 62.0%; Score 438.4; DB 17; Length 456;
Best Local Similarity 97.6%; Pred. No. 1.7e-95;

```
Matches 445; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 7 GCCTCCCTGTCGACCGGAGACCATCCCTTCAGGACCACTCGAGGGCACCGTGAACGGC 66
    |||||
Db 1 GCCTCCCTGTCGACCGGAGACCATCCCTTCAGGACCACTCGAGGGCACCGTGAACGGC 60
    |||||
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCAACCCCTTCGAGGGCAACCCAGAGATG 126
    |||||
Db 61 CACTGCTTCAAGTGCATCGGCAAGGCGAGGCAACCCCTTCGAGGGCAACCCAGAGATG 120
    |||||
QY 127 AAGATCGAGGTGATCGAGGCGCGCCCTCGCCCTTCCACATCTGTCACCTCC 186
    |||||
Db 121 AAGATCGAGGTGATCGAGGCGCGCCCTCGCCCTTCCACATCTGTCACCTCC 180
    |||||
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGCACTCCCGACTACTTCAAG 246
    |||||
Db 181 TGCATGTACGGCTCCAAAGACCTTCAATCAAGTACGTGTCCGCACTCCCGACTACTTCAAG 240
    |||||
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGGACCACTACGAGGACGGCGGCTTC 306
    |||||
Db 241 CAGTCCCTTCCCGAGGGCTTCACCTGGAGCGGACCACTACGAGGACGGCGGCTTC 300
    |||||
QY 307 CTGACCGGCCCAACAGGACACCTCCCTGGACCGGCACTGCTGGTGTACAAAGTGAAGATC 366
    |||||
Db 301 CTGACCGGCCCAACAGGACACCTCCCTGGACCGGCACTGCTGGTGTACAAAGTGAAGATC 360
    |||||
QY 367 CTGGGCAACAATCTCCCGCGCAGCGCCCTGTGATGAGAACAAAGCGCGCTGGAG 426
    |||||
Db 361 CTGGGTAACTTCCCGCGCAGCGCCCTGTGATGAGAACAAAGCTCGCGCTGGAG 420
    |||||
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGACGGCTG 462
    |||||
Db 421 CCCGAAACCGAGATCGTGTACGAGGTGACGGCTG 456
    |||||
```

RESULT 12

```
US-10-155-809-1
; Sequence 1, Application US/10155809
; Publication No. US20030092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Chudakov, Dmitry
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10/155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 767
; TYPE: DNA
; ORGANISM: anthozoa
US-10-155-809-1
```

Query Match 61.8%; Score 436.8; DB 15; Length 767;
Best Local Similarity 76.7%; Pred. No. 4.2e-95;
Matches 534; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

```
QY 7 GCCTCCCTGTCGACCGGAGACCATCCCTTCAGGACCACTCGAGGGCACCGTGAACGGC 66
    |||||
Db 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGAGCACCATTGAAGGACGGTTAATGGC 131
    |||||
```

67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGGAAACCCCTCGAGGGCAACCCAGGAGATG 126
Db |||
132 CACTACTTCAAGTGCACGGAAAGAGAGGGCAACCCATTTGAGGGTACGAGGAATG 191
Qy |||
127 AAGATCGAGGTGATCGAGGGGGGGCCCTCGCTTCGCTTCCACATCTCTGTCACCTCC 186
Db |||
192 AAGATAGAGGTGATCGAGGAGGGTCCATTCGCTTCCACATTTGCTTCCACATTTTGTCAACGAGT 251
Qy |||
187 TGCATGTACGGCTCCAAAGGCTTCAATCAAGTACGTGTCGGGCATCCCGGACTACTTCAAG 246
Db |||
252 TGTATGTACGGTATGAGACCTTCAATCAAGTATGTGTGAGGATTCCTGACTACTTCAAG 311
Qy |||
247 CAGTCCCTCCCGAGGGCTTCACTCGGAGCGCACCACTACGAGGAGCGGCTTC 306
Db |||
312 CAGTCTTCCCTGAAGGTTTACCTTGGGAAAGAGGAGGCAACCCATTTGAGGGTACGAGGAATG 371
Qy |||
307 CTGACCGCCACAGGACACCTCCCTCGAGCGGCACTGCTGTGTACAGGTGAAGATC 366
Db |||
372 CTTACAGCTCATCAGGACAAAGCTAGATGGAGATTGGCTTCGTTTACAGGTCAAGATT 431
Qy |||
367 CTGGGCAACACTTCCCGCGGAGCGCCCGGTGATGAGAAACAAAGCGCGCCCTCGGAG 426
Db |||
432 CTTGGTAATATTTCTGCTGATGGCCCGGTGATGAGAAACAAAGCGGAGATGGAG 491
Qy |||
427 CCCTCCACAGGATGCTGTACGAGGTGAGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 486
Db |||
492 CCAGCCACCGAGATAGTTTATGAAAGTTGACGGTCTCTGCTGGACAGTCTTTCATGGCC 551
Qy |||
547 AAGCCCGCTCCGCTTGAAGATGCGCGGTTCACCTTCGAGGACCAACCGCATCGAGATC 606
Db |||
612 AAACGAGTGTGAGAGGCGAGTGTACAGCAGTACGAGCGCGGCGGCGCTAC 671
Qy |||
607 CTGAGGAGGTGAGAGGCGAGTGTACAGCAGTACGAGCGCGGCGGCGCTAC 666
Db |||
672 ATGAGGAAGTTGAGAAAGGCAAGTGTATTAACAGTACGAGGAGGAGGAGGAGGAGTAC 731

RESULT 13

US-10-809-3
; Sequence 3, Application US/10155809
; Publication No. US20030092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Chudakov, Dmitzy
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Their Use
; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 767

; TYPE: DNA
; ORGANISM: anthozoa
US-10-155-809-3

Query Match 61.6%; Score 435.2; DB 15; Length 767;

Best Local Similarity 76.6%; Pred. No. 1e-94; Mismatches 163; Indels 0; Gaps 0;

Matches 533; Conservative 0;

Qy 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCTCGAGGGCAACCGTGAACGGC 66
Db 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTTGAAGGAGCGTTAATGGC 131
Qy 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGGCAACCCCTCGAGGAGGAGGAGGAGGAGT 126
Db 132 CACTACTTCAAGTGTACAGGAAAGAGGAGGAGGCAACCCATTTGAGGGTACGAGGAATG 191
Qy 127 AAGATCGAGGTGATCGAGGGGGGGCCCTCGCTTCCACATCTCTGTCCACCTCC 186
Db 192 AAGATAGAGGTGATCGAGGAGGTCCATTTGCCCTTCCACATTTTGTCAACGAGT 251
Qy 187 TGCATGTACGGCTCCAAAGGCTTCAATCAAGTACGTGTCGGGCATCCCGGACTACTTCAAG 246
Db 252 TGTATGTACGGTATGAGACCTTCAATCAAGTATGTGTGAGGATTCCTGACTACTTCAAG 311
Qy 247 CAGTCCCTCCCGAGGGCTTCACTCGGAGCGCACCACTACGAGGAGCGGCTTC 306
Db 312 CAGTCTTCCCTGAAGGTTTACCTTGGGAAAGAGGAGGCAACCCATTTGAGGGTACGAGGAT 371
Qy 307 CTGACCGCCACAGGACACCTCCCTCGAGCGGCACTGCTGTGTACAGGTGAAGATC 366
Db 372 CTTACAGCTCATCAGGACAAAGCTAGATGGAGATTGGCTTCGTTTACAGGTCAAGATT 431
Qy 367 CTGGGCAACACTTCCCGCGGAGCGCCCGGTGATGAGAAACAAAGCGCGCCCTCGGAG 426
Db 432 CTTGGTAATATTTCTGCTGATGGCCCGGTGATGAGAAACAAAGCGGAGATGGAG 491
Qy 427 CCCTCCACAGGATGCTGTACGAGGTGAGCGGCTGCTGCGGCGGAGTCCCTGATGGCC 486
Db 492 CCAGCCACCGAGATAGTTTATGAAAGTTGACGGTCTCTGCTGGACAGTCTTTCATGGCC 551
Qy 487 CTGAGGTGCGCGGGGTGCGGCTGACCTGACCTGACACCACTGACACCACTACCGTCCAG 546
Db 552 CTTAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
Qy 547 AAGCCCGCTCCGCTTGAAGATGCGCGGTTCACCTTCGAGGACCAACCGCATCGAGATC 606
Db 612 AAACGAGTGTGAGAGGCGAGTGTACAGCAGTACGAGCGCGGCGGCGCTAC 671
Qy 607 CTGAGGAGGTGAGAGGCGAGTGTACAGCAGTACGAGCGCGGCGGCGCTAC 666
Db 672 ATGAGGAAGTTGAGAAAGGCAAGTGTATTAACAGTACGAGGAGGAGGAGGAGGAGTAC 731
Qy 667 TGCAGCGCGCGCCCTCCAAAGCTGGGCGCAACTGA 702
Db 732 TGTGATGCTGCTCCATCCAGCTTGGACATTAATA 767

RESULT 14

US-10-155-809-5
; Sequence 5, Application US/10155809
; Publication No. US20030092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Chudakov, Dmitzy
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Their Use
; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176

; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 767
; TYPE: DNA
; ORGANISM: anthozoa
US-10-155-809-5

Query Match 51.3%; Score 433.6; DB 15; Length 767;
Best Local Similarity 76.4%; Pred. No. 2.4e-94;
Matches 532; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 7 GCCTCCTGCTGACCGAGACCATGCTCCCTCAGGACCACTGAGGCGACCGTGAACGCG 66
DB 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAGACGACCATTCAGGGACGGTTAATGC 131
QY 67 CACTACTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCAACCCAGGAGATG 126
DB 132 CACTACTCAAGTGTACAGGAAAGGAGAGGGCAACCCATTTGAGGGTACGAGGAAATG 191
QY 127 AGATGAGGTGATCAGGAGGGGCGCCCTCGCCCTTCGCTTCCACATCTGTCACCTCC 186
DB 192 AAGATAGAGTCAATCAGAGAGTCCATTTGCCATTTGCCATTTCCACATTTTGTCAAGGAT 251
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
DB 252 TGTATGTACGGTATGAGACCTTCATCAAGTATGTGTGAGGAATTCCTGACTACTTCAAG 311
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGCAACCACTACAGGACCGCGCTTC 306
DB 312 CAGTCTTTCCTGAAAGTCTTACTTGGGAAAGAACCAACCTACAGGATGGAGCTTT 371
QY 307 CTGACCGCCCAACAGGACACTCCCTGCGAGCGACTGCTGGTGATCAAGAGTGAAGATC 366
DB 372 CTTCAGCTCATCAGGACACAGCTAGATGAGATTGCTCGTTTACAAGTCAAGATT 431
QY 367 CTGGCAACAACTTCCCGCGCAGCGCCCTGTGATGAGAGAAAGGCGCGCGCTGGAG 426
DB 432 CTTGGTAATAATTTCTGCTGATGCGCCGCTGATGACAGAAACAGGAGGAG 491
QY 427 CCCTCCACCGAGATCGTGTACGAGTGGACGGGTGCTGCGGCGCAGTCCCTGATGGCC 486
DB 492 CCAGGACACGAGATAGTTTGAAGTGTACGGTGTCTGCTGGGACAGTCTTTGATGGCC 551
QY 487 CTGGAGTGGCCCGCGGTGCGCACTGACCTGACCTGACCTGACACCACTACCGCTCCAAG 546
DB 552 CTTAAGTGCCCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 611
QY 547 AAGCGCGCTCCCGCTGAGATGCGCGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
DB 612 AATCCAGCTAGTGCCTTGAAGTGGCAGGATTTTATTTGAAGATTACCGCATCGAGATA 671
QY 607 CTGGAGAGGTGAGAGGAGGCAAGTGTCTAAGCAGTACGAGCGCGCGCTGGCGCGCTAC 666
DB 672 ATGAGGAACTTGAGAAAGGCAAGTGTCTAATACAGTACAGCAGCAGTGGGCGAGTAC 731
QY 667 TGGGACCGCGCGCTTCAAGCTGGGCGCAACTGA 702
DB 732 TGTGATGCTCTCCATCCAGCTTGGACATACTAA 767

RESULT 15

US-09-976-673-11

; Sequence 11, Application US/09976673
; Patent No. US20020160473A1
; GENERAL INFORMATION:

; APPLICANT: Lukyanov, Sergey
; APPLICANT: Pradkov, Arcady
; APPLICANT: Labas, Yulii
; APPLICANT: Matz, Mikhail
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Gurskaya, Nadezda
; TITLE OF INVENTION: FAR RED SHIFTED FLUORESCENT PROTEINS
; FILE REFERENCE: CLON-028WO
; CURRENT APPLICATION NUMBER: US/09/976,673
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,018
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: 60,306,131
; PRIOR FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 687
; TYPE: DNA
; ORGANISM: heteractis crispa
US-09-976-673-11

Query Match 59.1%; Score 417.8; DB 9; Length 687;
Best Local Similarity 79.1%; Pred. No. 1.5e-90;
Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CCTGTGACCGAGACCATGCTCCCTTCAGGACCACTGAGGGCACCGTGAACGCGCACTA 71
DB 12 CCTGCTGAAGGAGAGTATGCGCATCAAGTGTACATGAGGGCACCGTGAACGCGCACTA 71
QY 72 CTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCAACCCAGGAGATGAAGAT 131
DB 72 CTTCAAGTGCAGGGCGGAGGGCGGCAACCCCTTCGCGCGGACCCAGAGCATGAGAAAT 131
QY 132 CGAGTGTATCGAGGGCGGCGCCCTCGCCCTTCGCTTCCACATCTGCTGCCACTCTGCTCAT 191
DB 132 CCAGTGTACCGAGGGCGGCGCCCTCGCCCTTCGCTTCCGACATCTGCGCCCTGCTGCGGA 191
QY 192 GTACGGTCTCAAGGCTTCATCAAGTGTGCGGCAATCCCGCACTACTTCAAGCAGTC 251
DB 192 GTACGGCAGCAGGACCTTCGTGTCACCAACCGCGGAGATCCCGCACTTCTTCAAGCAGAG 251
QY 252 CTTCCCGGAGGCTTCACCTGGAGAGAACCCACCACTTACAGGACGGCGCATCTCTGAC 311
DB 252 CTTCCCGGAGGCTTCACCTGGAGAGAACCCACCACTTACAGGACGGCGCATCTCTGAC 311
QY 312 CGCCCAACAGGACACTCCCTGGACGGCACTGCTGTGTACAAGTGAAGATCTCTGGG 371
DB 312 CGCCCAACAGGACACTCCCTGGAGGCACTGCTGTGTACAAGTGAAGTGAAGTGCACGG 371
QY 372 CAAACACTTCCCGCGGAGCGGCGCGCTGATGAGAGCAAGGCGCGCGCTGGAGCGCTC 431
DB 372 CACCAACTTCCCGCGGAGCGGCGCGCTGATGAGAGCAAGGCGCGCGCTGGAGCGCGAG 431
QY 432 CACCGAGATCGGTACAGAGTCCCGGCTTCGCGCGGCGAGTCCCTGTGATGGCGCTCGA 491
DB 432 CACCGAGTGTGTATCCCGGAGAGCGGCTGCTGTGCGCGGAGAACGATGATGGCGCTGAA 491
QY 492 GTGCGCGCGGCTGCGCACTGACCTGCACTGCAACCACTACCGCTCCAGAGAGCC 551
DB 492 G---GTGCGCGGACCGGCACTGATCTGCCACCACTACACCACTACCGAGCAAGAGGC 548
QY 552 GCCTCCCGCTGAGAGTCCCGGCTTCGAGGAGCAACCGCATCTGAGATCGAGATCTCTGGA 611
DB 549 CGTGCGCGCTTACCATGCGCGCTTCCACTTACCGAGATCCGGCTCCAGATGCTGG 608
QY 612 GGAGGTGGAGAGGGCAAGTGTCTAAGCAGTACGAGCGCGCGCTGGCGCGCTACTGGA 671
DB 609 GAA---GAAGAGGAGCAGGATCTTCGAGCTGTACGAGGCGGAGCGTGGCGCGGTACAGGA 665
QY 672 C 672
DB 666 C 666

Search completed: August 3, 2004, 19:01:43
Job time : 516.043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 11:02:24 ; Search time 3167.97 Seconds
(without alignments)
6664.376 Million cell updates/sec

Title: US-10-081-864-21
Perfect score: 707
Sequence: 1 ggaatcgctccctgctgac.....ctgggccaactgaagctt 707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513299 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.2	14.3	550	13	BQ551001
2	73.6	10.4	562	14	CA937115
3	72.2	10.2	625	12	BG906607
4	71.8	10.2	749	14	CF489659

5	71	10.0	852	14	CD438780
6	69.8	9.9	549	12	BI943903
7	69.4	9.8	606	12	BI942400
8	69.4	9.8	986	12	BM561582
9	69.4	9.8	990	10	BE216934
10	69	9.8	593	29	CG332384
11	68.8	9.7	697	13	BQ526888
12	67.6	9.6	496	13	BQ578547
13	67.6	9.6	631	14	CA754991
14	67.6	9.6	695	14	CD902570
15	67.4	9.5	869	14	CK159167
16	67.2	9.5	925	29	CNS0091P
17	66.6	9.4	545	12	BI974926
18	66.2	9.4	799	14	CB656196
19	66	9.3	753	29	CG231715
20	66	9.3	835	29	CG51563
21	66	9.3	842	29	CG273302
22	66	9.3	2299	11	AY106831
23	65.4	9.3	715	10	BE519855
24	65.4	9.3	728	14	CF484651
25	65.2	9.2	982	13	BX415111
26	65	9.2	473	14	CD866846
27	64.8	9.2	672	14	CA501436
28	64.8	9.2	674	13	BQ255271
29	64.8	9.2	822	29	CG301217
30	64.8	9.2	852	29	CG311347
31	64.8	9.2	929	29	CC719024
32	64.4	9.1	624	6	AL811149
33	64.2	9.1	538	12	BM178969
34	64.2	9.1	644	14	CF486437
35	64.2	9.1	647	14	CD233835
36	63.8	9.0	967	13	BX428434
37	63.8	9.0	1079	14	CK210801
38	63.6	9.0	546	12	BI780772
39	63.6	9.0	622	12	BG839331
40	63.6	9.0	778	14	CF480713
41	63.6	9.0	792	14	CF488222
42	63.6	9.0	1147	14	CK209786
43	63.4	9.0	822	29	CG337308
44	63.4	9.0	877	29	CG337294
45	63.4	9.0	885	29	CC727899

ALIGNMENTS

RESULT 1
BQ551001
LOCUS H4006G01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H4006G01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BQ551001
VERSION BQ551001
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 550)
AUTHORS VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G.,
Martin,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T.,
Kargul,G.J., Luo,A.G., Kello,J., Hide,W. and Ko,M.S.H.
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse
JOURNAL CDNA clone set
MEDLINE Genome Res. 12 (12), 1999-2003 (2002)
PUBMED 22354164
COMMENT Other ESTs: H4006G01-3
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@nigun.grc.nia.nih.gov

550 bp mRNA linear EST 20-JUN-2002
Mus musculus cDNA clone

CD438780 EL01N0517
BI943903 ea52309.Y
BI942400 sb80112.Y
BM561582 AGENCOURT
BE216934 EST0477.I
CG332384 OGVDD58TV
BU626888 UI-H-FTO
BQ578547 WHE0306.E
CA754991 BR030010L
CD902570 G356.107G
CK159167 FGAS04055
AL033013 Drosophila
BI974926 sal74g12.
CB656196 OSJNEC10E
CG231715 OGVG062TH
CC651863 OGVBR82TC
CG273302 OGVBO38TV
AY106831 Zea mays
BE519855 HV_CEB002
CF484651 POL1_26.D
BX415111 BX415111
CD866846 G118.103H
CA501436 WHE4033.F
BQ255271 WHE2868.B
CG301217 OGVFR89TH
CG311347 OGVBC17TC
CC719024 OGGAC14TV
AL811149 Triticum
BM178969 sa761f08.
CF486437 POL1_37.E
CD233835 S81_4.F02
BX428434 BX428434
CK210801 FGAS02262
BI780772 EBes01.SQ
BG839331 Gm01_14e0
CF480713 POL1_67.A
CF488222 POL1_48.B
CK209786 FGAS02156
CG337308 OGVFR58TV
CG337294 OGVFR58TV
CC727899 OGKAO61TH

This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
 Plate: H4006 row: G column: 01
 Seq primer: -21M13 Reverse
 High quality sequence stop: 550
 POLYA=No.

FEATURES

Location/Qualifiers

source

```

1..550
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="niaEST:H4006G01-5"
/db_xref="taxon:10090"
/clone="H4006G01"
/sex="mixed"
/dev_stages="mixed"
/lab_host="DH10B"
/clone_lib="DH10B"
/Note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 7,407 clones from more than 20 cDNA libraries."

```

ORIGIN

```

Query Match      14.3%; Score 101.2; DB 13; Length 550;
Best Local Similarity 57.4%; Pred. No. 0.00023;
Matches 225; Conservative 0; Mismatches 158; Indels 9; Gaps 2;

QY 13 CTGTGACCGGAGACCATGCGCCCTTCAGGACACCACTGAGGGCAGCGGCAAGCCACTAC 72
DB 159 CTGTTCCCGGGTGGTGGCCATCTCTGTGAGCTGACGGCGAGCTAAACGGCCACAG 218
QY 73 TTCAAGTGACCGGCAAGGGGAGGCAACCCCTCGAGGGACCCAGGAGATGAAGATC 132
DB 219 TTCAAGCTGTCCGCGAGGGGAGGCGATGCCACCTACGCGAAGCTGACCTGAAGTTC 278
QY 133 GAGGTGATCGAGGGCGGCGCCCTCGCCCTTCACATCTCTGTCACCTCTGTCATG 192
DB 279 ATCTGACACAC---CGGAGAGCTCCCGTGGCCCTGCGCCACCTCTGTCGACCCCTGACC 335
QY 193 TACGGCTCCAGGCGCTTCATCAAGTACGTGTCGGCAT-----CCCGACTACTTCAAG 246
DB 336 TACGGCGTGAGTCTTCAGCGGTACCCGACACGTGAAGCAGCAGCACTTCTTCAAG 395
QY 247 CAGTCCCTCCCGAGGCTTCACTGGAGGAGGACCACTACGAGGAGCGGCGCTTC 306
DB 396 TCCGCCATGCGCGAAGGCTAGCTCAGGAGCGCACCATCTTCTTCAAGGACGAGCAAC 455
QY 307 CTGACCGCGCCACAGGACACCTCCCTCGAGCGGACCTGCTGTGTACAGGTGAAGATC 366
DB 456 TACAGACCCCGCGGAGGTGAAGTTCGAGGGGACACCTCTGTGAACCCGATCGAGCTG 515
QY 367 CTGGGCAACAACTTCCCGCGGAGCGGCCGCT 398
DB 516 AAGGGCATCGACTTCAAGGAGGAGCGGCAACAT 547

```

RESULT 2

CA937115

LOCUS

DEFINITION sav29605.y1 Gm-c1048 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1048-6754 5' similar to SW:ER1_HEVER Q39963 ETHYLENE-INDUCIBLE PROTEIN HEVER. ; mRNA sequence.

CA937115

VERSION

CA937115.1

KEYWORDS

SOURCE

ORGANISM

Glycine max (soybean)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE

1 (bases 1 to 562)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V.,

Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, J., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 407.

FEATURES

Location/Qualifiers

source

```

1..562
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3947"
/clone="SOYBEAN CLONE ID: Gm-c1048-6754"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stages="1 week old"
/lab_host="DH10B"
/clone_lib="Gm-c1048"
/Note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The Clark NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from whole seedlings of 1 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the phagescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). The library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."

```

ORIGIN

```

Query Match      10.4%; Score 73.6; DB 14; Length 562;
Best Local Similarity 47.0%; Pred. No. 1.2;
Matches 263; Conservative 0; Mismatches 294; Indels 3; Gaps 1;

QY 36 CAGGACCAACCATCGAGGCAACCGTGAACGGCCACTACTTCAAGTGCACCGGCAAGGGCGGA 95
DB 1 CACGATCCCGCTGATGGCCAAAGGCCCGCATCGGCCACTTCTCGAAGCCGACATCTCTCGA 60
QY 96 GGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAGTGTAGGCGGCGCCCT 155
DB 61 AGCCATCGGCATCGACTACTGTCGAGAGAGGAGGTCTACCTCCCGGACGAGCGCAA 120
QY 156 GCCCTTCGCCCTCCACATCTCTGTCCACCTCTCTGTATGTACGGTCTCAAGGCCCTTCATCAA 215
DB 121 CCACATCAACAAACACAACTCCGCATCCCTCTCGTGTGGCTGCCGCAACCTCGCGGA 180
QY 216 GTACGTGTCCGGCATCCCGGACTACTTCAAGCATCTCTCCCGAGGGCTTCACCTGGGA 275
DB 181 GGCCCTTCGGCGCATCCCGGAGGCGCGGCATATCCGACCAAGGCGGAGGCGCGCAC 240
QY 276 GCGCACCACCACTTACGAGGACGCGGCTCTCTGACCGGCCACCCAGGACACCTCCCTCGGA 335
DB 241 CGGCAACATCATCGAGCGGCTTCGCCAGCTCCGCTCGTCTATGAGCGACATTAGGGTCT 300
QY 336 CGGCGACTGCTGTGTACAGGTGAAGATCTCTGGGCAACACTTCCCG---CCGACGG 392

```


/mol_type="mRNA"
 /cultivar="Brix623"
 /db_xref="taxon:4558"
 /clone="POL1_58_H06_A002"
 /lab_host="DH10B-T1 phage-resistant E. coli"
 /clone_lib="Pollen"
 /note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line Brix623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Query Match 10.2%; Score 71.8; DB 14; Length 749;
 Best Local Similarity 48.4%; Pred. No. 2.1;
 Matches 233; Conservative 0; Mismatches 242; Indels 6; Gaps 1;

QY 9 CTCCTGCTGACCGAGACCATGCCCTTCAGGACACCATCGAGGCAACCGTGAACGCCA 68
 DB |||||
 QY 207 CTCGCTCGCGCGGTGGCTCGGCTCCCGAGTACTCCGACTCCGCCAAGGTGGCTA 266
 DB |||||
 QY 69 CTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGCAACCCAGAGATGAA 128
 DB |||||
 QY 267 CGTCTCCAGGACATGACCTCGGGATCGTCTTCCGAGGCGGCAAGGAGAGGT 326
 DB |||||
 QY 129 GATGAGGTGATCAGGCGCGCCCTCGCTTCCGCTTCCACATCTCTGCCACTCTCTG 188
 DB |||||
 QY 327 GGTGCTCGTGAAGAGCGAGCGCTGGCTTCTCCCTTGGCGTCTGTGATGGTGGCA 386
 DB |||||
 QY 189 CATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAGCA 248
 DB |||||
 QY 387 CAAGCCGCCACCGGTGACGACCTGACGCTCTCTTCTCGGCGACACCTCCAAAGG 446
 DB |||||
 QY 249 GTCCCTCCCGAGGGCTCACCTGGAGGCGCACCAACCATAGAGGACGGCGGCTTCT 308
 DB |||||
 QY 447 CCACAGCCAGCGGAGTTCAACAATTCCAGCCGCTCTGGCATCTCAAGACTCCAT 506
 DB |||||
 QY 309 GACCGCCACACGAGACACTCTCCCTGGAGCGGACTGCTGTGTACAA-----GTTGAA 362
 DB |||||
 QY 507 CGCGTCCGAGCGGCTCGCCCAAGGACCGGAGGGCATGCGCTCAACTGCCCTTGAGGC 566
 DB |||||
 QY 363 GATCTGGGCAACAACTTCCCGCGGAGGCGCCCGTGTATGAGNAACAAGGCGGCGCTG 422
 DB |||||
 QY 567 GCCGTGACGTGACATATCCCGCGAGGCGCGTGGTGTCTCAACACCGTGAACCT 626
 DB |||||
 QY 423 GGAGCCCTCCACCGAGATCGTGTACGAGTGGACGGGCTGTGCGCGGCCAGTCCCTGAT 482
 DB |||||
 QY 627 GCCCTTGTCAAGACGTGGGCTGGCGCGCACTGTCGGCATCAACGGCACTCCAT 686
 DB |||||
 QY 483 G 483
 DB 687 G 687

RESULT 5
 CD438780 852 bp mRNA linear EST 03-JUN-2003
 LOCUS
 DEFINITION E010N0517C04.b Endospem_5 Zea mays cDNA, mRNA sequence.
 ACCESSION CD438780
 VERSION CD438780.1 GI:31354423
 KEYWORDS EST.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

REFERENCE

AUTHORS

clade: Panicoidae; Andropogoneae; Zea.
 1. (bases 1 to 852)
 Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,B. and Messing,J.

TITLE

Sequencing of the maize endospem ESTs

JOURNAL

Unpublished (2002)

COMMENT

Contact: Lai, Jinsheng
 Dr. Joachim Messing's lab
 Wakman Institute, Rutgers University
 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
 Tel: 732-445-3801
 Fax: 732-445-5735
 Email: jlai@wakman.rutgers.edu
 Seq primer: T3.

FEATURES

source

1..852
 /organism="Zea mays"
 /mol_type="mRNA"
 /cultivar="W22"
 /db_xref="taxon:4577"
 /tissue_type="Endospem of 7-23DAP"
 /clone_lib="Endospem 5"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 10.0%; Score 71; DB 14; Length 852;
 Best Local Similarity 47.5%; Pred. No. 2.8;
 Matches 277; Conservative 0; Mismatches 300; Indels 6; Gaps 2;

QY 2 GATCGCCCTCCCTGTGACCGAGACCATGCCCTTCAGGACACCATCGAGGCAACCGTGA 61
 DB |||||
 QY 259 GGTGGACTCCACCCCGCGCTGTCCAGGCTCGCGGGCTCTCTCGCTTCACCGGCC 318
 DB |||||
 QY 62 ACGGCGACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCAGGCGCAACCGG 121
 DB |||||
 QY 319 GCCTGGCGAGCCCTCGCGGCTGTGAGCGCTACGGCTTCGGCCCGTCCACACCTCT 378
 DB |||||
 QY 122 AGATGAAGATGAGGTGATCGAGGCGCGCCCTGCTTCCGCTTTCACATTCCTGTCCA 181
 DB |||||
 QY 379 CCATCTCTCTCCCGCGCACCCCGCGCTCTCTCTCTCGCGCGAGCACCGCGTGTCTCC 438
 DB |||||
 QY 182 CCTCTGTGATGACGGTCCAG--GCCTTCATCAAGTACGTGTCCGGCATCCCGACT 238
 DB |||||
 QY 439 CGCCCTGGGTGACCCCTCCGCTCCGCGTCCGCGCGCGGAGGAGCTCGGACCCCGACA 498
 DB |||||
 QY 239 ACTTCAAGCAGTCTCCCTCCCGAGGCTTCACTGGAGCGGCAACCACTACAGGAGC 298
 DB |||||
 QY 499 CCACCGCGCAGGTCCACCCCGGCTCTTCAACAAGGCGCTCTCGCGCGTCCGAGCGC 558
 DB |||||
 QY 299 GCGGCTTCTGACCGCCACACAGACACTCCCTGGAGCGGAGCTCCCTGTGTGTACAGG 358
 DB |||||
 QY 559 AGGTCTCATTTGGCGAGGTGGAGCGGTGTGCTGGCGCGGAGCGGCGCGTCCGCGCG 618
 DB |||||
 QY 359 TGAAGATCTCTGGGCAACAACCTTCCCGCGCGAGCGGCGCTGATGAGAACAGGCGCGCC 418
 DB |||||
 QY 619 TCGTGTCAAGGGCGCGACCGCGTGTGAGCGCGCGCGCTGTCTGTCTCGCTCGGCC 678
 DB |||||
 QY 419 GCTGGAGCCCTCCACCGAGATCTGTGTACGAGGTGGAGCGGTGTCTGCGCGCGAGTCC 478
 DB |||||
 QY 679 CGTGGTCCGCGCGCTCCAGGTGTTCAGCGAGGTTCAGCTGTCCGGGTTCAGGCGCG 738
 DB |||||
 QY 479 TGATGCCCTGGAGTGGCGCGGCTCGCCACTGACCTGCACCTGCACACCACTACCTACC 538
 DB |||||
 QY 739 ACAGCATCTGTCTCGGCGCGCGGAGCGCGA---GAAGTCAACGCGACTGCTCTTCC 795
 DB |||||
 QY 539 GTTCAAGAAGCCCGCTCCGCGCTGAAGATGCCCGGTTCOA 581
 DB |||||
 QY 796 TCAGTACACGCGCGAGCGCGCGCGCAAGATGCTCGACCCGA 838
 DB |||||

RESULT 6

B1943903


```

QY 431 CCACCGAGATGCTAGACGGTGAACGGCGTG-----CTGGCGGGCCAGTCCC 478
Db 484 CTGGGACCAACAGAAATCGCGGCTGGACGGCGAGTTCCGGCAAGTCTCGTGT 543
QY 479 TGATGGCCCTGGAGTGCCTGGCGGCTGCACCTGACCTGCCACCTGSCACACCACTACC 538
Db 544 GGAGGACTTGGACAAACCCAGGTGCTGGCCCTGGATCCACCAAGGGTACATTAAT 603
QY 539 GCTCCAGAAAGCCCGCTCCGCCCTGAAGATGCCCGGCTTCCATTCTGAGGACCAACCGCA 598
Db 604 GGACCGAGTGGCGGGCGAAGCCGAGGATCGTGGGGCTTTCATGGACGGGACCAACTGCA 663
QY 599 TCGAGATCTCGAGGAGTGGAGAGGCGAAGTCTACAGACAGTACGAGGCGCGCTGG 658
Db 664 TGACGCTGGTGGACAAAGTGGCGGGCCCAACACCTTACCATTAAGTACGCTGACCAAGC 723
QY 659 GCCGCTACTGCGAGCGCGCCCTCCCAAGCTGGGCGCA 695
Db 724 GCCTCTACTGACCGACTGGACACCAACATGATCGA 760

RESULT 9
BE216934 990 bp mRNA linear EST 03-JUL-2000
LOCUS EST0477 Triticum aestivum Lambda Zap Triticum aestivum cDNA clone
DEFINITION JAI 5B H05_T3 5', mRNA sequence.
ACCESSION BE216934
VERSION BE216934.1 GI:9904620
KEYWORDS EST.
ORGANISM Triticum aestivum (bread wheat)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE Anderson, J.M., Williams, C.E. and Goodwin, S.B.
AUTHORS Analysis of an EST database reveals a probable CP2 resistance gene
TITLE homolog in wheat
JOURNAL Unpublished (2000)
COMMENT Contact: Anderson, J.M.
Crop Production & Pest Control Research Unit
USDA-ARS
1150 Lilly Hall, West Lafayette, IN 47907, USA
Tel: 765-494-5565
Fax: 765-496-2926
Email: janderson@purdue.edu
Seq primer: T3
High quality sequence stop: 990.
FEATURES
    source
        1..990
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /strain="P29"
            /db_xref="taxon:4565"
            /clone="JAI_5B_H05_T3"
            /tissue_type="leaf"
            /dev_stage="9 day old seedlings"
            /clone_lib="Triticum aestivum Lambda Zap"

ORIGIN
    Query Match 9.8%; Score 69.4; DB 10; Length 990;
    Best Local Similarity 46.9%; Pred. No. 4.7; Indels 3; Gaps 1;
    Matches 251; Conservative 0; Mismatches 281;

QY 32 CTTTCAGGACCAACCATCGAGGGCAACCGTGAACGGCCACTACTTCAAGTCACCGGCAAGG 91
Db 211 CCGTCACCATCCAGTCATGSCCAAGGCCCGCATCGGCGACTTCGTGAGGCCCATTC 270
QY 92 GCGAGGGCAACCCCTCGAGGGCAACCGAGGATGAAGATCGAGGTGATCGAGGGCGGCC 151
Db 271 TCGAGTCCATCGGGCTCGACTACGTGAGCAGGAGCGAGGTCTCAGCTCGCGACGACG 330
QY 152 CCCTGCGCTTCGCTTCCACATCTGTCCACCTCTCTCATGTACGGCTCCAGGCTCA 211

```

```

Db 331 CCCACCATCAACAAAGCACACTTCCGCTCCCTTCTGCTGCGGTGCGCCAACTGG 390
QY 212 TCAAGTAGTGTTCGGGCATCCCGGACTACTTCAAGCAGTCCCTCCCGAGGGCTTCACT 271
Db 391 GCGAGCCCTCGCGCGCATCCGCGAGGGCGCGCATGATCGCACCAAGGCGAGGCGG 450
QY 272 GGGAGGACCAACCACTACGAGGACGGGGGTTCCTACCGCCACCAGGACACTTCCC 331
Db 451 GCACCGGCAATGATCGAGGCGCTCCGCCACGTCGCTCCGTCTCATGGCGGACGTGCGTG 510
QY 332 TGGACGGGACTCGCTGTGTACAAAGGTGAAGATCTGGGCAACAACATTCCTCCCG--CG 388
Db 511 CCTCGGAGCATGACGACGAGGATTCACCTATGCCAGAGCATCGCGGACCAT 570
QY 389 ACGGCCCCGTGATGAGAAACAGGCGCGCTGGAGCCCTTCCACCGAGATCGTGTACG 448
Db 571 ACGACTCTGTCATGACAGCAAGCAGCTCGGCGCGCTCGCTCCGCTCCGATTCGCGCGCG 630
QY 449 AGGTGAGCGGTGCTGCGGCGCCAGTCCCTGATGCGCCCTCATGACCTCGCTGCGGCGCTG 508
Db 631 GCGGGGTGCCACGCCGCGGCGATGCGGCTCATGACCTCGCTGCGGCGCTGCTGTCT 690
QY 509 ACCTGACCTGCCACCTGCACACCACTACCGCTCCAAAGAGCCCGCTCCGCGCT 563
Db 691 TCGTCGGCTCCGCTGTCTTCAAGAGCGGGACCCCGCGCGCTGCGCGCGCAT 745

RESULT 10
CG332384 593 bp DNA linear GSS 26-AUG-2003
LOCUS OGVD58TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBma0509120,
DEFINITION genomic survey sequence.
ACCESSION CG332384
VERSION CG332384.1 GI:34249650
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
AUTHORS Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OGVD58TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: Sheared ends.
FEATURES
    Location/Qualifiers
        1..593
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBma0509120"
            /clone_lib="ZM 0.7 1.5 KB"
            /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

ORIGIN
    Query Match 9.8%; Score 69; DB 29; Length 593;
    Best Local Similarity 46.2%; Pred. No. 4.9;
    Matches 265; Conservative 0; Mismatches 305; Indels 3; Gaps 1;
    QY 2 GATCGCTCCCTGCTGACCGGAGACCATGCTCCCTTCAGGACCACTGAGGCGCACCTGTA 61

```

Db	23	GGTGGACTCCACCCCGCGCTCTCCAGCTCGCGGGGCTCTCTTGGCGTGACCGCC	82
Qy	62	ACGGCCACTACTTCAAGTGCAACCGGACGAGGCGAACCCCTCTGAGGGCACCCAGG	121
Db	83	GCTGGCGACGCGCTCGCGCGCGGACGCTACGGGCTTCGCGCGCTCCACACCTCT	142
Qy	122	AGATCAAGATCGAGTGATCGAGGCGCGCCCTCGCTTCCGCTTCACATCTCTGCCA	181
Db	143	CGGTCTCTCCCCCGCACCCGCGCTCTCTCTCGCGCGCGACCCGCTGCTCCCG	202
Qy	182	CCTCTGATGATCGGCTTCAAGGCTTTCATCAAGTACGTGTCCGGCATCCCGACTACT	241
Db	203	CCTGGGTGACCCCTCGCGCTCGCGCGCGCGCGGAGGAGCTCGGACCCCGACACCA	262
Qy	242	TCAAGCATGCTCTCCCGAGGCTTCACTGGGAGCGGACACCACTACGAGACGGCG	301
Db	263	CGCGCAGGTTCACCCGCGCTCTTCAACAAAGGCGCTCTCGCGCGCTCGGAGCGGAG	322
Qy	302	GCTTCTCTGACCGCCACACGACACCTCCCTGGAGCGGACTGCTGTGTACAAAGTGA	361
Db	323	TGCTCATCGGAGGTGAGCGGCTGCGCGCTGGAGCGCGGCTCGCGCGCTCG	382
Qy	362	AGATCTCTGGCAACAACTTCCCGCGGACGCGCGCGGTGATGATGACAAAGCGCGCT	421
Db	383	TGGTCAAGGGCGCGACGCGCTGTGGAGCGGACGCGCGCTGCTGCGCTCGCGCGCT	442
Qy	422	GGAGCCCTTCCACGAGATCTGTACGAGGTGAGCGGCTGCTGCGCGCGCGCTGCA	481
Db	443	GTCGCGCGCGCTCGAGGTGTGACGAGGTGTGACGAGGTGTGACGAGGTGTGACG	502
Qy	482	TGGCCCTTGAGTGTCCCGCGCGCTGCGCACTGACCTGACCTGACCTGACCAACCA	541
Db	503	GCATGCTCTCGCGCGCGCGGAGCGCA--GAAGGTCAAGCGCGCTGCTCTCTCA	559
Qy	542	CAAGAAAGCGCGCTCGCGCTGAGATGCGCG	574
Db	560	GCTACACCGGAGCGCGCGCGCAAGATGCTCG	592
RESULT 11			
LOCUS	BU626888	697 bp	mRNA
DEFINITION	UI-H-F70-bhn-d-01-0-UI.s1 NCI CGAP_F70 Homo sapiens CDNA clone	linear	EST 23-SEP-2002
ACCESSION	BU626888		
VERSION	1		
KEYWORDS	EST		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
TITLE	Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Robeff-Pamela, U of I CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu Seq primer: M13 FORWARD POLYA=Yes.		
FEATURES	Location/Qualifiers		
source	1..697		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="UI-H-F70-bhn-d-01-0-UI"		
ORIGIN			
Query Match	9.7%;	Score 68.8;	DB 13; Length 697;
Best Local Similarity	48.5%;	Pred. No. 5.4;	
Matches	224;	Conservative 0;	Mismatches 232; Indels 6; Gaps 1;
Qy	232	CCGCTACTTCAAGCAGTCCCTCCCGAGGCTTCACTGGGAGCGGACCAACCACTAC	291
Db	653	CACGACTTCTTCAAGTCCGCAATGCCGGAAGTAACGTCCAGGAGCGGACCATCTTC	594
Qy	292	GAGGACGGCGGCTCTCTGACCGCCACCAAGACCGCGCGAGGTGAAGTTCGAGGGG	351
Db	593	AAGGACGAGCGCACTACAGACCGCGCGAGGTGAAGTTCGAGGGGACACCTCGTG	534
Qy	352	TACAAAGGTGAAGTCTGGGCAACAACTTCCCGCGGAGCGGCGGCTGATGAGACAG	411
Db	533	AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCGCAACATCCTGGG	474
Qy	412	GCGGCGCGTGGAGCGCTCCACCGAGATCGTGTACGAGGTGGAGCGGCTGTGCGCG	471
Db	473	CTGGAGTACACTACACAGCCACACGCTCTATATCATGGCCGACACAGAGAACG	414
Qy	472	CAGTCCCTGATGGCGCTGAGGTGCCCGGGTGCACCTGACCTGACCTGACCACTG	531
Db	413	ATCAAGGTGAAGTCTCAAGATTCGCGCAC-----AACATCGAGGACGCGGCTG	360
Qy	532	ACCTACCGCTCCAGAGCGCGCTCCCGCTGAGATGCCCGGCTTCCACTTCGAGGAC	591
Db	359	GCGGACCACTACAGAGAACACCCCATCGCGGAGCGGCGGCTGCTGCTGCGGAC	300
Qy	592	CACCGCATCGAGATCTCGAGGAGGTGGAGAGGACCAAGTGTACAAAGGTACAGG	651
Db	299	CACCTACCTGAGCACCCAGTCCGCGCTGAGCAAGACCCCAAGAGAGGCGGATCA	240
Qy	652	GCGTGGCGGCTACTGCGAGCGCGCGCTCCAGCTGGCG	693
Db	239	GTCCTGCTGGAGTCTGTGACCGCGCGCGGATCACTCTCGCG	198
RESULT 12			
LOCUS	BQ578547	496 bp	mRNA
DEFINITION	WHE0306_E02_I04ZS Wheat unstressed seedling shoot cDNA library	linear	EST 19-JUN-2002
ACCESSION	BQ578547		
VERSION	1		
KEYWORDS	EST		
SOURCE	Triticum aestivum (bread wheat)		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.

REFERENCE

1 (bases 1 to 496)
 Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D.,
 Han, P.S., Hsia, C.C., Kang, Y., Iazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.L. and Tong, J.C.

TITLE

The structure and function of the expressed portion of the wheat
 genomes - Etiolated shoot cDNA library

JOURNAL

COMMENT

Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105955773
 Fax: 5105955818

Email: candersn@pw.usda.gov

Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20

Seq primer: SK primer.

FEATURES

source

1..496

Location/Qualifiers

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="WHE0306_E02_I04"

/tissue_type="Etiolated shoot"

/dev_stage="Five day old seedling"

/lab_host="E. coli SOLR"

/clone_lib="Wheat unstressed seedling shoot cDNA library"

/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Seeds were
 surface-sterilized, germinated and grown aseptically in
 the dark at room temperature on filter paper with water,
 nystatin and cefotaxime in covered crystallization
 dishes. Shoots were harvested. The tissue, total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the TJ Close lab (Choi, Close, Fenton) at the
 University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."

ORIGIN

Query Match

Best Local Similarity 48.3%; Pred. No. 7.5; Length 496;

Matches 221; Conservative 0; Mismatches 234; Indels 3; Gaps 1;

138 GATCGAGGGCGCCCTCGCCCTTCCACATCCTGTCCACCTCTCGCATGTACGG 197

8 GCTCGCGGACGCGCCACACATCAACAGCACAACTCCGGTCCCTTGTCTGGG 67

198 CTCGAGGCTTATCAAGTACGTGTCCGCGATCCCGGACTACTTCAAGCAGTCCCTCC 257

68 CTCGCGGACCTGGGCGAGGCCCTCCGCGCATCCGCGAGGGCGCGCATGATCGGTAC 127

258 CGAGGGGTTTCACTGGGAGCGCACCACTACGAGGACGGGGTTCCTGACCGCCCA 317

128 CAGGGCGAGGCGGACCGGATGTATCGAGGCGCTCGGACCGTCCGCTCGTAT 187

318 CAGGACACCTCCTGTGACGGGCGTCCGTGTGTGTCACAGGTGAAGATCTGGGCAACAA 377

188 GGGCGAGCTCGTGTCCCTCCGAGCATGGACGATGACGAGGTATTACCTATGCAAGAG 247

378 TTTCGCGCGG---ACGGCCCGGTGTCAGACAAAGCGCGCGCTGGAGCCCTCCAC 434

248 CATCGCGGACCGTACGACCTCTCATGACAGACCAAGCAGCTCGTTCGCTCGCGTCT 307

435 CGAGATGCTGTACGAGTGGACGGCGTGTCTGCGCGCCAGTCTCCCTGATGCCCTGGAGTG 494

308 CCAGTTCCGCGCGGGGTGCGCCACGCGCGCGATGCGGCCCTCATGATGACGCTCG 367

495 CCGCGCGGTGCGGCACCTGACCTGCCACCTGACACACACCTACCGTCCCAAGAGCCCG 554

368 CTGCGACGGTGTCTTCTGTCGGCTCGGTGTCTTCAAGAGCGCGACCCGCGCGCGGC 427

555 CTCGCGCCCTGAGATGCCCGCTTCCACTTCGAGGACC 592

428 CGCGCCCATCTGTCAGGCGGTCAACCCACTACAGGACC 465

RESULT 13

CA754991

LOCUS

DEFINITION

BR030010000_PLATE_D08_60_073.ab1 OA Oryza sativa (japonica

cultivar-group) cDNA clone BR030010000_PLATE_D08_60_073.ab1 similar

to unknown protein [Oryza sativa (japonica cultivar-group)], mRNA

sequence.

CA754991

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 631)

Bohner, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M.,

Palacio, C., Scarra, G., McCallough, A., Michalowski, C.B.,

Functional Genomics of Plant Stress Tolerance

Unpublished (2000)

Contact: Mark Fredrickson

Department of Plant Biology

University of Illinois

1201 W. Gregory Dr., Urbana, IL 61801, USA

Tel: 2172855473

Email: bohnertlab@life.uiuc.edu.

Location/Qualifiers

1..631

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="BR030010000_PLATE_D08_60_073.ab1"

/tissue_type="roots"

/dev_stage="3-4 weeks"

/clone_lib="OA"

/note="19 h 200mM NaCl"

Query Match

Best Local Similarity 47.7%; Pred. No. 7.7;

Matches 196; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

56 CCACCTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTCGAGGACCCAGGAGAT 125

8 CCCCACCCCATCGTCCGCGACCGGAGGTGGCGGACCAACCGCGCGCGCGCTGGCGAC 67

126 GAAATCATGAGTGTATCGAGGGCGGCCCTTCCCTTTCCTTCCACATCTCTGTCCACCTC 185

58 GGCACAGGTATCATCCCGTCTTCGCGTCCGCGCTCGGCCCCACCTCGCCAGGTCCA 127

186 CTGATGTAGGGCTTCAAGGCTTTCATCAAGTAGTGTGTCCGGCATCCCGGACTACTTCAA 245

128 CGGCACCGGCGACGCCCAACCTCCACCATCTCCATCCATCCCGCCNACGTCCA 187

246 GCAGTCCCTCCCGAGGGCTTTCACCTGGGAGCGACACACCTACGAGGACGGCGTT 305

188 GATCGCGCGCGCGAGCTCCCGGAGCACCCTCCCGCGGCGCCATGCCGGGACAGTT 247

306 CTTGACCGCCCAACGAGCACCTCCCTGGACGGGACTGCTGTGTGTACAGGTGAAGAT 365

248 CCAGGAGCGCGCAAGGACATCTCTGTGTGTCTGTGGGATCTCTTCGCGCTCGGCTG 307

366 CTTGGGCAACACTTCCCGCGGAGGCGCCGTGTATGCAGAACAGGCGCGCGCTGGGA 425

Wed Aug 4 09:01:06 2004

[illegible]

Search completed: August 3, 2004, 15:48:03
Job time : 3171.97 secs

Blanks Shot USSPTD